

Qy 1 STAPPVHN 9
Db 210 STAPPVHN 218

RESULT 2

A;Cross-references: EMBL:M31823; PID:9181542; PIDN:AAA35757.1; PID:9181543
R;Masuzawa, Y.; Miyauchi, T.; Hamanoue, M.; Ando, S.; Shimazu, H.
J. Biochem. 112, 609-615, 1992

A;Title: A novel core protein as well as polymorphic epithelial mucin carry peanut agglu-

A;Reference number: JX0235; MUID:1478919
A;Accession: PX005
A;Molecule type: mRNA
A;Residues: 998-1011, 'ES', 1014-1017, 1018-1032, 'T', 1034-1037, 1038-1057 <MAS>
A;Experimental source: Gastric carcinoma cell
R;Zribi-Licht, S.; Baruch, A.; Eliray-Stein, O.; Keydar, I.; Wreschner, D.H.
FBBS Lett. 356, 130-136, 1994

A;Title: Tyrosine phosphorylation of the MUC1 breast cancer membrane proteins
A;Reference number: SS1026; PMID:988707
A;Contents: annotation
A;Note: undetermined tyrosine residues in the carboxyl-terminal non-repetitive region are
C;Comment: This protein is length polymorphic. Individuals may have between 21 and 125 c-
partial repeats. The repeat shown is defined by SmaI nucleotide sites.
C;Comment: Serine and threonine residues in the tandem repeat domain are extensively gly-
C;Comment: For an alternative splice form without a tandem repeat domain, see PIR:S48146
C;Genetics:
A;Gene: GDB:MUC1, PUM
A;Cross-references: GDB:120705; OMIM:158340
A;Map Position: 1q21-1q23
A;Introns: 20/1: 62/3; 1165/3; 1184/2; 1230/1; 1270/3; 1320/3
C;Superfamily: Polymorphic epithelial mucin
C;Keywords: alternative splicing; duplication; glycoprotein; phosphoprotein; polymorphis-
F:1-134/Predicted: mucin 1 precursor splice form A #status predicted <PRE>
F:1-62/Region: mucin 1 amino-terminal non-repetitive
F:1-23/Domain: signal sequence #link PREB #status predicted <SIGA>
F:1-19/2-3/Domain: signal sequence #link PREB #status predicted <SIGB>
F:1-19,29-134/Region: mucin 1 precursor, splice form B #status predicted <PREB>
F:1-19,29-212/103-1344/Product: mucin 1 precursor, epithelial tumor antigen splice form
F:138-1017/Region: 20-residue repeats (GATAPPAGVNTAPDTTPAP)
F:1143-1344/Region: mucin 1 carboxyl-terminal non-repetitive
F:1245-1272/Domain: transmembrane #status predicted <TRM>
F:1046,1064,1118,1144,1122/Binding site: carbohydrate #status predicted <TRM>
F:1213/Binding site: phosphate (Tyr) (covalent) #status predicted
F:1213/Binding site: phosphate (Tyr) (covalent)

Query Match 100.0% Score 49; DB:1; Length: 1344;
Best Local Similarity 100.0%; Pred. No. 0.53; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 STAPPVHN 9
Db 1039 STAPPVHN 1047

RESULT 3

A;Cross-references: GB:J052739; GB:J05288; PID:9182126; PIDN:AAA35806.1; PID:9182129; GB:
A;Experimental source: splice form B
A;Note: Genbank entries HOMEPIS1A1 and HOMEPIS1A2 present only the amino-and carboxyl-en-
A;Accession: B35175
A;Molecule type: mRNA
A;Residues: 1-19,29-952,1033-1344 <LG2>
A;Cross-references: GB:J052739; GB:J05288; PID:9182126; PIDN:AAA35806.1; PID:9182129; GB:
A;Experimental source: splice form B
A;Note: Genbank entries HOMEPIS1B1 and HOMEPIS1B2 present only the amino-and carboxyl-en-
A;Title: R.G.; Gendler, S.J.; Lancaster, C.A.; Taylor-Papadimitriou, J.; Duhig, T.; Peat, N.; Burchell
J. Biol. Chem. 265, 15286-15293, 1990
A;Title: Molecular cloning and expression of human tumor-associated polymorphic epitheli-
A;Reference number: A35886; MUID:90268715; PMID:169589
A;Accession: A35886
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-19,29-952,1033-1344 <GEN>
A;Cross-references: GB:J05581; PID:9188869; PIDN:AAA59876.1; PID:9188870
A;Note: Human entry HUMNUCAB includes one copy of the tandemly repeated sequence
R;Lan, M.S.; Batra, S.K.; Qi, W.N.; Metzgar, R.S.; Hollingsworth, M.A.
J. Biol. Chem. 265, 15294-15299, 1990
A;Title: Cloning and sequencing of a human pancreatic tumor mucin cDNA.
A;Reference number: A35887; MUID:90268716; PMID:2394722
A;Accession: A35887
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-19,29-1109, 'S', 1111-1339, 'A', 1341-1344 <LAN>
A;Cross-references: GB:J05582; PID:9189582; PIDN:AAA60019.1; PID:9189599
A;Note: Genbank entry HUMANMU contains four fewer copies of the tandemly repeated seque-
R;Wreschner, D.H.; Harneveni, M.; Tearfatty, I.; Smorodinsky, N.; Horev, J.; Zaretsky, J.
Bur. J. Biochem. 189, 463-473, 1990
A;Title: Human epithelial tumor antigen cDNA sequences. Differential splicing may genera-
A;Reference number: S10571; MUID:90276413; PMID:2351132
A;Accession: S10572
A;Molecule type: mRNA
A;Residues: 1-19,29-155, 'P', 157-175, 'P', 177-182, 'A', 184-212, 1033-1344 <WRE>
A;Cross-references: EMBL:X52229; NID:937053
R;Wreschner, D.H.
Submitted to the EMBL Data Library, March 1990
A;Reference number: S40293
A;Accession: S40293
A;Molecule type: mRNA
A;Residues: 1-19,29-155, 'P', 157-175, 'P', 177-182, 'A', 184-212, 1033-1037, 'A', 1039-1344 <WR2>
A;Cross-references: EMBL:X52229; NID:937053; PIDN:CHA36478.1; PID:937054
R;Abe, M.; Siddiqui, J.; Kufe, D.
Biochem. Biophys. Res. Commun. 165, 644-649, 1989
A;Title: Sequence analysis of the 5' region of the human DF3 breast carcinoma-associated
A;Reference number: A36735; MUID:90088473; PMID:2591151
A;Accession: A36735
A;Molecule type: mRNA
A;Residues: 1-142, 'Q', 144-162, 'Q', 164-168 <ABE>

RESULT 4
AF2555 hypothetical protein all8037 [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC7120ga
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriuchi, N.; Nakazaki, N.; Shimpoo, S.; Sugimoto, M.; Takazawa, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena Reference number: AB1007; MUID:21595285; PMID:11759440
A;Accession: AF2555
A;Status: preliminary
A;Residues: 1-590 <KUR>
A;Cross-references: GB:AP003603; PID:BAB77367.1; PID:gi17134810; GSPDB:GN00182
C;Genetics:
A;Gene: all8037
A;Experimental source: strain PCC 7120
A;Genome: plasmid

Query Match 83.7%; Score 41; DB 2; Length 590;
Best Local Similarity 87.5%; Pred. No. 7;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TAPPVHN 9
Db 9 TAPPVHN 16

RESULT 5
S38177 SSY7 protein homolog - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein YKR098c
C;Species: Saccharomyces cerevisiae
C;Accession: S38177
C;Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 06-Feb-1998
R;Gallion, L.; Dujon, B.
Submitted to the Protein Sequence Database, March 1994
A;Reference number: S38177
A;Accession: S38177
A;Molecule type: DNA
A;Residues: 1-717 <GA>
A;Cross-references: EMBL:Z28323; NID:9486592; PID:gi486593; MIPS:YKR098C
A;Experimental source: strain S288C
C;Genetics:
A;Gene: SGD:YBRP11
A;Cross-references: SGD:S0001806; MIPS:YKR098C
A;Map Position: 11R

Query Match 77.6%; Score 38; DB 2; Length 717;
Best Local Similarity 66.7%; Pred. No. 32;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 STAPPVHN 9
Db 231 ATAPPVHSL 239

RESULT 6
T39233 Inositol polyphosphate phosphatase - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
R;Churcher, C.M.; Gentles, S.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
Submitted to the EMBL Data Library, August 1997
A;Reference number: 221837
A;Accession: T39233
A;Status: preliminary; translated from GB/BMBL/DBBJ

A;Molecule type: DNA
A;Residues: 1-1183 <CHU>
A;Cross-references: EMBL:Z298763; PIDN:AB11494.1; GSPDB:GN00066; SPDB:SPAC9G1.10C
C;Genetics:
A;Gene: SPDB:SPAC9G1.10C
A;Map Position: 1

Query Match 75.5%; Score 37; DB 2; Length 1183;
Best Local Similarity 77.8%; Pred. No. 83;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 STAPPVHN 9
Db 253 STAPPVHN 261

RESULT 7
T13016 hypothetical protein F8L21.40 - Arabidopsis thaliana (mouse-ear cress)
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Accession: T13016
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 22-Oct-1999
R;Bevan, M.; Peters, S.A.; van Staveren, M.; Dirks, W.; Stiekema, W.; Bancroft, I.; Mew submitted to the Protein Sequence Database, July 1999
A;Reference number: T13016
A;Molecule type: DNA
A;Accession: T13016
A;Residues: 1-329 <BEV>
A;Cross-references: EMBL:AL096882; GSPDB:GN00062; ATSP:F8L21.40
A;Experimental source: cultivar Columbia; BAC clone F8L21
C;Genetics:

Query Match 73.5%; Score 36; DB 2; Length 329;
Best Local Similarity 55.6%; Pred. No. 33;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 STAPPVHN 9
Db 307 SS1PPVHN 315

RESULT 8
A70851 probable trxB2 protein - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Accession: A70851
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A;Authors: Soares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Accession number: A70500; MUID:98295987; PMID:9634320
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Accession: A70851
A;Residues: 1-335 <COL>
A;Cross-references: GB:AL021426; GB:AL123456; PIDN:CAA16226.1; PID:gi280869
C;Genetics:
A;Gene: trxB2
C;Superfamily: thioredoxin reductase; thioredoxin reductase homology <trxB2>
F;11-315/Domain: thioredoxin reductase homology <trxB2>

Query Match 73.5%; Score 36; DB 2; Length 335;
Best Local Similarity 85.7%; Pred. No. 33;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TAPPVHN 8

Db 2 TAPPVHD 8

RESULT 9
 S54775 Query Match 71.4%; Score 35; DB 2; Length 154;
 cell size regulation protein RCS1 - yeast (Saccharomyces cerevisiae) C;Species: Caenorhabditis elegans
 N;Alternate names: AFT1 protein;protein Q32666; protein YGL071W
 C;Species: Saccharomyces cerevisiae
 C;Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 29-Oct-1999
 C;Accession: S54775; S54874; S12915; S64078
 R;Iamaguchi, Iwai, Y.; Dancic, A.; Klausner, R.D.
 EMBL: U_14..1231..1239_1995
 A;Title: AFT1: a mediator of iron regulated transcriptional control in Saccharomyces cerevisiae
 A;Reference number: S54775; MUID:95237204; PMID:7720713
 A;Accession: S54775
 A;Molecule type: mRNA
 A;Residues: 1-690 <YAM>
 A;Cross-references: EMBL:Z48004; PIDN:CAA88044.1; PID:97582778
 R;Aldea, M.; Casas, C.; Gallego, C.; Espinet, C.; Herrero, E.
 Submitted to the EMBL Data Library, January 1994
 A;Reference number: S54874
 A;Accession: S54874
 A;Molecule type: DNA
 A;Residues: 1-7..H', 9-70, 'S', 72-135, 'T', 137-235, 'N', 237-415, 'G', 417-506, 'S', 508-537, 'P', 90 <ALD>
 A;Cross-references: EMBL:X77413; PIDN:J9840840; PIDN:CAA54586.1; PID:9840881
 R;Gil, R.; Zueco, J.; Sentandreu, R.; Herrero, E.
 Yeast, 7, 1-14, 1991
 A;Title: RCS1, a gene involved in controlling cell size in Saccharomyces cerevisiae.
 A;Reference number: S12915; MUID:91213643; PMID:2021081
 A;Accession: S12915
 A;Molecule type: DNA
 A;Residues: 370-377, 'L', 379-415, 'G', 417-467, 'T', 469-506, 'S', 508-537, 'L', 539-567, 'P', 569-
 A;Cross-references: EMBL:X57046; PIDN:94294; PIDN:CAA7215.1; PID:94295
 R;Rieger, M.; Mueller-Auer, S.; Brueckner, M.; Schaeffer, M.
 Submitted to the Protein Sequence Database, May 1996
 A;Reference number: S64078
 A;Molecule type: DNA
 A;Residues: 1-690 <RIE>
 A;Cross-references: EMBL:Z72593; PIDN:CAA96775.1; PID:e243959; PID:9132258
 A;Experimental source: strain S288C
 C;Genetics:
 A;Gene: SGD:RCS1; AFT1
 A;Cross-references: SGD:S0003039; MIPS:YGL071W
 A;Map position: 7L
 C;Function:
 A;Description: control of iron uptake

Query Match 73.5%; Score 36; DB 2; Length 690;
 Best Local Similarity 75.0%; Pred. No. 72;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 STAPPVHN 8
 Db 497 SSAPPVH 504

RESULT 10
 S33873 Query Match 71.4%; Score 35; DB 2; Length 347;
 hypothetical protein 154 - Streptomyces ambofaciens C;Species: Streptomyces ambofaciens
 C;Date: 19-May-1995 #sequence_revision 26-May-1995 #text_change 26-May-2000
 C;Accession: S33873; S33421
 R;Hagege, J.; Pernodet, J.L.; Friedmann, A.; Guerineau, M.
 Mol. Microbiol. 10, 799..812, 1993
 A;Title: Mode and origin of replication of pSAM2, a conjugative integrating element of S
 A;Reference number: S33873; MUID:9502051; PMID:934842
 A;Accession: S33873
 A;Molecule type: DNA
 A;Residues: 1-154 <HAG>
 A;Cross-references: EMBL:Z19590; PIDN:CAA79638.1; PID:9298045

Query Match 71.4%; Score 35; DB 2; Length 347;
 Best Local Similarity 55.6%; Pred. No. 54;
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 STAPPVHN 9
 Db 324 STGPSIHNI 332

RESULT 13
T46907 hypothetical protein DKFZp761G2023.1 - human
C;Species: Homo sapiens (man)
C;Date: 17-Mar-2000 #sequence_revision 17-Mar-2000
C;Accession: T46907
R;Bloecher, H.; Boecker, M.; Brands, P.; Meves, H.W.; Weil, B.; Wiemann, S.
A;Reference number: Z24137
A;Accession: DKFZp761G2023.1
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-363 <AAA>
A;Cross-references: EMBL:AL136570
A;Experimental source: adult amygdala; clone DKFZp761G2023
C;Genetics:
A;Note: DKFZp761G2023.1

Query Match	71.4%	Score 35;	DB 2;	Length 363;
Best Local Similarity	100.0%	Pred. No. 56;	0;	Gaps 0;
Matches	6;	Conservative	0;	Mismatches 0;
Qy	2 TAPPVH 7 	Db	337 TAPPVH 342	

A;Accession: S14554
A;Molecule type: DNA
A;Residues: DVCCSPLIQYSV' 261-335 <SCH>
A;Cross-references: EMBL:X51819; NID:95249; PID:CAA36119.1; PID:g255726
R;Hankley, C.; Leibham, D.; Perry, M.
submitted to the EMBL Data Library, January 1991
A;Description: Regulated binding specificity of Xenopus OCT-1, a maternal transcription factor
A;Accession: S08673
A;Molecule type: DNA
A;Residues: DVCCSPLIQYSV' 261-335 <SCH>
A;Cross-references: EMBL:X51819; NID:95249; PID:CAA36119.1; PID:g255726
R;Hankley, C.; Leibham, D.; Perry, M.
submitted to the EMBL Data Library, January 1991
A;Description: Regulated binding specificity of Xenopus OCT-1, a maternal transcription factor
A;Accession: S14554
A;Molecule type: DNA
A;Residues: 1-77, 'L' 9-760 <SM2>
A;Cross-references: EMBL:X51819; NID:95249; PID:CAA36119.1; PID:g255726
R;Schiltz, J.G.; Baarends, W.M.; Petersen-Maduro, J.; Destre, O.H.J.
submitted to the EMBL Data Library, February 1990
A;Reference number: S08673
A;Accession: S14554
A;Molecule type: mRNA
A;Residues: 1-156, 'S' 158-367,370-760 <HIN>
A;Cross-references: EMBL:X51765; NID:954951; PID:CAA40454.1; PID:g64952
R;Baltzinger, M.; Stiegler, P.; Remy, P.
Nucleic Acids Res. 18, 6131, 1990
A;Title: Cloning and sequencing of POU boxes expressed in Xenopus laevis neurula embryos
A;Accession: S12179; NID:91045083; PMID:2235499
A;Molecule type: mRNA
A;Residues: 320-435 <BAL>
A;Cross-references: EMBL:X54683; NID:954947; PID:g930281
C;Genetics:
A;Gene: Oct-1
A;Introns: 320/3
C;SuperFamily: transcription factor Oct-1; homeobox homology; POU domain homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;301-368/Domain: POU domain homology <POU>
F;398-454/Domain: homeobox homology <HOX>
A;Accession: E844682
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-653 <STO>
A;Cross-references: GB:AE002093; NID:94803957; PID:ADD29829.1; GSPDB:GR00139
C;Genetics:
A;Gene: At2g28240
A;Map position: 2

RESULT 14
E844682 hypothetical protein At2g28240 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001
C;Accession: E844682
R;Lin, X.; Kahl, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.; euss, D.; Nieman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J. A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-653 <STO>
A;Cross-references: GB:AE002093; NID:94803957; PID:ADD29829.1; GSPDB:GR00139
C;Genetics:
A;Gene: At2g28240
A;Map position: 2

Query Match	71.4%	Score 35;	DB 1;	Length 760;
Best Local Similarity	77.8%	Pred. No. 1.2e-02;	Matches 7;	Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy	1 STAPPVHNV 9 :	Db	529 STAPPVSSV 537	

Search completed: July 28, 2004, 18:11:08
Job time : 17 secs

RESULT 15
S07896 transcription factor Oct-1-32 - African clawed frog
N;Alternate names: maternal transcription factor; pou homeobox protein oct-1.32
C;Species: Xenopus laevis (African clawed frog)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 15-Oct-1999
R;Smith, D.P.; Old, R.W.
Nucleic Acids Res. 18, 39, 1990
A;Title: Nucleotide sequence of Xenopus laevis Oct-1 cDNA.
A;Reference number: S07896; MUID:90221827; PMID:2326173
A;Accession: S07896
A;Molecule type: mRNA
A;Residues: 1-760 <SM1>
A;Cross-references: EMBL:X17190; NID:964942; PID:g671665

- Best Local Similarity 100.0%; Pred. No. 0.091; Mismatches 0; Indels 0; Gaps 0;
- Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
- Qy 1 STAPPVENV 9
Db 170 STAPPVHN 178
- RESULT 2**
- MUC1 HUMAN STANDARD; PRT; 1255 AA;
AC P15941; P13931; P15942; P17626; Q14128; Q14876; Q16437; Q16442;
AC Q16615; Q9BX4A; Q9UE75; Q9uQ11; Q9Y4U2; [7]
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-OCT-1990 (Rel. 14, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
- Mucin 1 precursor (MUC-1) (Polymorphic epithelial mucin) (PEM) (PEMT)
DE "Cloning and sequencing of a human pancreatic tumor mucin cDNA.";
DE (Episialin) (Tumor-associated mucin) (Carcinoma-associated mucin)
DE (Tumor-associated epithelial membrane antigen) (EMA) (H2AG) (Peanut-reactive urinary mucin) (PM) (Breast carcinoma-associated antigen)
DE (CD227 antigen).
GN MUC1.
RN Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
NCBI_TAXID=9606; [1]
RN SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Pancreas; MEDLINE=90368716; PubMed=2394722;
RA Lan M.S., Batra S.K., Qi W.-N., Metzgar R.S., Hollingsworth M.A.;
RT "Cloning and sequencing of a human pancreatic tumor mucin cDNA.";
RL J. Biol. Chem. 265:15294-15299 (1990).
RN SEQUENCE FROM N.A. (ISOFORM 1 AND 2).
RC TISSUE=Breast; MEDLINE=90202794; PubMed=2318825;
RA Lightenberg M.J.L., Vos H.L., Gennissen A.M.C., Hilkens J.;
RT "Episialin, a carcinoma-associated mucin, is generated by a polymorphic gene encoding splice variants with alternative amino termini.";
RT Peat N., Burchell J., Pemberton L., Malani E.-N., Wilson D.;
RT "Molecular cloning and expression of the human tumor-associated polymorphic epithelial mucin.";
RL J. Biol. Chem. 265:15573-15578 (1990).
RN SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Breast carcinoma; MEDLINE=90368715; PubMed=1697589;
RA Gandler S.J., Lancaster C.A., Taylor-Papadimitriou J., Duhig T.,
RA Peat N., Burchell J., Pemberton L., Malani E.-N., Wilson D.;
RT "Molecular cloning and expression of the human polymorphic epithelial mucin gene: an expressed VNTR unit.";
RL Biochem. Biophys. Res. Commun. 173:1019-1023 (1990).
RN SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Breast carcinoma; MEDLINE=907524; PubMed=2263039;
RA Lancaster C.A., Peat N., Duhig T., Wilson D.,
RA Taylor-Papadimitriou J., Gandler S.J.;
RA "Structure and expression of the human polymorphic epithelial mucin gene: an expressed VNTR unit.";
RL Biochem. Biophys. Res. Commun. 173:1019-1023 (1990).
RN SEQUENCE FROM N.A. (ISOFORM 5).
RC TISSUE=Breast carcinoma; MEDLINE=9076413; PubMed=2351132;
RA Wreschner D.H., Hareveni M., Tsarfaty I., Smorodinsky N., Horev J.,
RA Zaretsky J., Kotkes P., Weiss M., Lathe R., Dion A., Keydar I.;
RA "Human epithelial tumor antigen cDNA sequences. Differential splicing may generate multiple protein forms.";
RL Eur. J. Biochem. 189:463-473 (1990).
RN SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Breast carcinoma; MEDLINE=90276414; PubMed=2112460;
RA Hareveni M., Tsarfaty I., Zaretsky J., Kotkes P., Horev J., Keydar I., Wreschner D.H.,
RA Zriban S., Weiss M., Green S., Lathe R., Keydar I., Wreschner D.H.,
RA "Mucin mRNA expression in lung adenocarcinoma cell lines and tissues.";
- RT "A transcribed gene, containing a variable number of tandem repeats, codes for a human epithelial tumor antigen. cDNA cloning, expression of the transcribed gene and over-expression in breast cancer tissue.";
RT "Isolation and characterization of an expressed hypervariable gene coding for a breast-cancer-associated antigen.";
RN SEQUENCE FROM N.A. (ISOFORM 1).
RP MEDLINE=9103045; PubMed=168329;
RA Tsarfaty I., Hareveni M., Horev J., Zaretsky J., Weiss M., Wreschner D.H.;
RA Jeltisch J.W., Garnier J.M., Lathe R., Keydar I., Wreschner D.H.;
RA "Characterization and molecular cloning of a novel MUC1 protein, devoid of tandem repeats, expressed in human breast cancer tissue.";
RL Eur. J. Biochem. 224:787-795 (1994).
RN [9]
RN SEQUENCE FROM N.A. (ISOFORMS 6; 7 AND 8).
RX RX MEDLINE=9735747; PubMed=9212226;
RA Zrhan-Licht S., Vos H.L., Baruch A., Elroy-Stein O., Sagiv D.,
RA Keydar I., Hilkens J., Wreschner D.H.;
RA Oosterkamp H.M., Scheiner L., Stefanova M.C., Lloyd K.O.,
RA Firstad C.L.;
RT "Comparison of MUC-1 mucin expression in epithelial and non-epithelial cancer cell lines and demonstration of a new short variant form (MUC-1/Z).";
RT Int. J. Cancer 72:87-94 (1997).
RN RN SEQUENCE FROM N.A. (ISOFORM 7).
RP RP SEQUENCE FROM N.A. (ISOFORM 7).
RA Zhang L.X., Li C.H.;
RA "Molecular cloning of an isoform of MUC1, MUC1LY.";
RT Submitted (FB-1999) to the EMBL/GenBank/DBJ databases.
RN [10]
RN SEQUENCE FROM N.A. (ISOFORM 9).
RP RP SEQUENCE FROM N.A. (ISOFORM 9).
RC TISSUE=Epithelial cancer; RA Zhang L.X., Li C.H., Sun L.Y., Yue W.;
RT "Cloning of a new potential secreted short variant form of MUC1 mucin in epithelial cancer cell line.";
RL Submitted (FB-2001) to the EMBL/GenBank/DBJ databases.
RN [11]
RN PARTIAL SEQUENCE FROM N.A. (ISOFORM 9).
RX RX MEDLINE=83330762; PubMed=3417635;
RA Gandler S.J., Taylor-Papadimitriou J., Duhig T., Rothbard J.,
RA Burchell J.;
RA "A highly immunogenic region of a human polymorphic epithelial mucin expressed by carcinomas is made up of tandem repeats.";
RL J. Biol. Chem. 263:12820-12823 (1988).
RN [12]
RN PARTIAL SEQUENCE FROM N.A. (ISOFORM 2).
RX RX MEDLINE=90089473; PubMed=2597151;
RA Gandler S.J., Taylor-Papadimitriou J., Duhig T., Rothbard J.,
RA Burchell J.;
RA "Sequence analysis of the 5' region of the human DF3 breast carcinoma-associated antigen gene.";
RT RT Biochem. Biophys. Res. Commun. 165:644-649 (1989).
RN [13]
RN SEQUENCE OF 1-160 FROM N.A. (ISOFORM 2).
RP RP SEQUENCE OF 1-109 FROM N.A. (ISOFORM 2).
RC TISSUE=Thyroid; RA Abe M., Siddiqui J., Kufe D.;
RC MEDLINE=96183746; PubMed=8608966;
RA Weiss M., Baruch A., Keydar I., Wreschner D.H.;
RA "Preoperative diagnosis of thyroid papillary carcinoma by reverse transcriptase polymerase chain reaction of the MUC1 gene.";
RT RT Int. J. Cancer 66:55-59 (1996).
RN [15]
RN SEQUENCE OF 1-89 FROM N.A. (ISOFORM 2).
RP RP TISSUE=Lung;
RX RX MEDLINE=96181716; PubMed=8604237;
RA Yu C.J., Yang P.C., She J.Y., Hong T.M., Yang S.C., Lee Y.C.,
RA Lee L.N., Luh K.T., Wu C.W.;
RT "Mucin mRNA expression in lung adenocarcinoma cell lines and tissues.";

RESULT 3	
RN	SEQUENCE OF 1-46 FROM N.A. (ISOFORMS 3 AND 4).
RC	TISSUE=Breast carcinoma;
RA	Buluwella L., Liu Q., Lugmani Y.A., Gomm J.J., Coombes R.C.; Submitted (OCT 1992) to the EMBL/GenBank/DBJ databases.
[17]	CARBOHYDRATE-LINKAGE SITES WITHIN THE REPEAT.
RP	MEDLINE=974460054; PubMed=9312074;
RA	Muller S., Goletz S., Packer N., Gooley A.A., Lawson A.M., Hanisch F.-G.;
RA	"Localization of O-glycosylation sites on glycopptide fragments from lactation-associated MUC1. All putative sites within the tandem repeat are Glycosylation targets <i>in vivo</i> .";
RT	J. Biol. Chem. 272:24780-24793(1997).
[18]	CARBOHYDRATE-LINKAGE SITES WITHIN THE REPEAT.
RP	MEDLINE=9303572; PubMed=1033415;
RA	Muller S., Alving K., Peter-Katalinic J., Zachara N., Gooley A.A., Hanisch F.-G.;
RA	"High density O-glycosylation on tandem repeat peptide from secretory MUC1 of T47D breast cancer cells.";
RL	J. Biol. Chem. 274:18165-18172(1999).
[19]	POLYMORPHISM WITHIN THE REPEAT.
RP	MEDLINE=1359466; PubMed=1030974;
RX	Egelmann K., Baldus S.E., Hanisch F.-G.;
RA	"Identification and topology of variant sequences within individual repeat domains of the human epithelial tumor mucin MUC1.";
RL	J. Biol. Chem. 276:27764-27769(2001).
RN	CHARACTERIZATION OF ISOFORM Y, AND MUTAGENESIS OF ASP-1116.
RP	MEDLINE=99211485; PubMed=10197628;
RA	Barich A., Hartmann M.-L., Yoeli M., Adereith Y., Greenstein S., Wreschner D.H.;
RA	"The breast cancer-associated MUC1 gene generates both a receptor and its cognate binding protein.";
RT	Cancer Res. 59:1552-1561(1999).
[21]	PARTIAL SEQUENCE, AND CHARACTERIZATION OF CLEAVAGE SITE.
RP	MEDLINE=1240104; PubMed=11341784;
RA	Parry S., Silverman H.S., McInermott K., Willis A., Hollingsworth M.A., Harris A.;
RA	"Identification of MUC1 proteolytic cleavage sites <i>in vivo</i> .";
RL	Biochem. Biophys. Res. Commun. 283:715-720(2001).
RN	CHARACTERIZATION.
RP	MEDLINE=21836452; PubMed=11847293;
RA	Wreschner D.H., McCucklin M.A., Williams S.J., Baruch A., Yoeli M., Ziv R., Okun L., Zaretsky J., Smorodinsky N., Keydar I., Neophytou P., Stacey M., Lin H.-H., Gordon S.;
RA	"Generation of ligand-receptor alliances by 'SEA' module-mediated cleavage of membrane-associated mucin proteins.";
RT	Protein Sci. 11:638-706(2002).
[23]	PHOSPHORYLATION.
RP	MEDLINE=95080414; PubMed=7988707;
RX	Zrihan-Licht S., Baruch A., Elroy-Stein O., Keydar I., Wreschner D.H.; RT
RT	"Tyrosine phosphorylation of the MUC1 breast cancer membrane proteins. Cytokine receptor-like molecules.";
RL	FEMS Lett. 356:130-136(1994).
CC	-!- FUNCTION: May play a role in adhesive functions and in cell-cell interactions, metastasis and signaling. May provide a protective
CC	interactions, metastasis and signaling. May provide a protective
Query Match	Score 49; DB 1; Length 1255;
Best Local Similarity	100.0%; Pred. No. 0.25+;
Matches 9;	Conservative 0; Mismatches 0; Indels 0; Gaps 0; Caps 0;
Qy	1 STAPPVHN 9 : : :
Db	231 ATAPPVHSL 239
RESULT 4	
TRXB	MYCTU STANDARD;
ID	TRXB MYCTU
AC	P5274; 053592; STANDARD;
DT	01-OCT-1996 (Rel. 34, Created)
DT	15-JUL-1999 (Rel. 38, Last sequence update)
DT	15-MAR-2004 (Rel. 43, Last annotation update)
DE	Thioredoxin reductase (EC 1.8.1.9) (TRXR) (TR).
GN	TRXB OR RV3913 OR MT4032 OR MT0208-04.
OS	Mycobacterium tuberculosis.
OC	Bacteria: Actinobacteria: Actinomycetidae: Actinomycetales:
OC	Corynebacteriales: Mycobacteriaceae; Mycobacterium.
NCBI_TaxID	1773;
NCBI_TaxID	1173;

RP	SEQUENCE FROM N.A.		FT	DISULFID	145	148	REDOX-ACTIVE (BY SIMILARITY).
RA	Wieles B., Phillip W., Drijfhout J.W., Offringa R.,		FT	NP_BIND	288	297	FAD (FLAVIN PART) (BY SIMILARITY).
RA	Ottenhoff T.H.M.,		FT	CONFFLICT	125	125	A -> R (IN REF. 1).
RT	"Sequence analysis and functional characterization of thioredoxin and thioredoxin reductase of Mycobacterium tuberculosis.";		FT	CONFFLICT	215	215	V -> C (IN REF. 1).
RT	submitted (MAR-1996) to the EMBL/GenBank/DDJB databases.		FT	CONFFLICT	228	228	V -> Y (IN REF. 1).
RN	[12]		SEQUENCE	335 AA;	35643 MW;	3D0DD581B6C187E2 CRC64;	
RP	SEQUENCE FROM N.A.		Query Match	73.5%	Score 36;	DB 1;	Length 335;
RX	STRAIN=H37RV;	PubMed=9634230;	Best Local Matches	85.7%	Pred. No. 17;		
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,		Local Similarity	85.7%;	Indels 0;	Gaps 0;	
RA	Gordon S.V., Eigemeier K., Gas S., Barry C.B. III, Tekala F.,		Conservative	1;	Mismatches 0;		
RA	Baddock K., Basham D., Brown D., Chillingworth T., Connor R.,						
RA	Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,						
RA	Hornsby T., Jagels K., Krogh A., McLean J., Moulie S., Murphy L.,						
RA	Oliver S., Osbourne J., Quail M.A., Rajandream M.A., Rogers J.,						
RA	Butter S., Seeger K., Skeleton S., Squares S., Squares R., Sulston J.E., Taylor K., Whitehead S., Barrell B.G.,						
RA	"Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.";						
RA	Nature 393:537-544(1998).						
RN	[13]						
RP	SEQUENCE FROM N.A.						
RX	STRAIN=CDC 1551 / Oshkosh;						
RX	MEDLINE=222065494 / Published=12218036;						
RA	Fleischmann R.D., Allard D., Eisen J.A., Carpenter L., White O., Peterson J., DeBoy R., Dodson R., Gwinn M., Halt D., Hickey E., Kolonay J.F., Nelson W.C., Omayam L.A., Ermolaeva M., Salzberg S.L., Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A., RA Bishai W., Jacobs W.R. JR., Venter J.C., Fraser C.M.; RA RWR whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains;"						
RA	J Bacteriol. 184:5479-5490(2002).						
CC	- - CATALYTIC ACTIVITY: Thioredoxin + NADP(+) = thioredoxin disulfide + NADPH.						
CC	- - COFACTOR: Binds 1 FAD per subunit (By similarity).						
CC	- - SUBUNIT: Homodimer (By similarity).						
CC	- - SUBUNITARY LOCATION: Cytoplasmic (By similarity).						
CC	- - MISCELLANEOUS: The active site is a redox-active disulfide bond.						
CC	- - MISCELLANEOUS: Belongs to class-II pyridine nucleotide-disulfide oxidoreductase family.						
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).						
CC							
DR	EMBL; X95798; CAA65070; 1; -						
DR	EMBL; AL021426; CAA16226; 1; -						
DR	EMBL; AE007194; AAC48397; 1; -						
DR	PRINTS; PR00419; ADXRDTASE.						
DR	INTERPRO; IPR000759; Adrindr reductase.						
DR	INTERPRO; IPR001327; FAD_py_redox.						
DR	INTERPRO; IPR001100; Pyr_redox.						
DR	INTERPRO; IPR008255; Pyr_redox2_AS.						
DR	INTERPRO; IPR00103; Pyridine_redox_2.						
DR	INTERPRO; IPR005982; Thiorredeox_reduct.						
DR	PFAM; PF00070; pyr_redox; 1.						
DR	PRINTS; PR00419; ADXRDTASE.						
DR	PRINTS; PR00368; FADPNR.						
DR	PRINTS; PR00411; PNDRTASEI.						
DR	TIGRFAMS; TIGR01294; TRX_reduct.						
DR	PROSITE; PS00573; PYRIDINE_REDox_2; 1.						
DR	Redox-active center; Oxidoreductase; NADP; Flavoprotein; FAD;						
KW	Complete proteome.						
FT	NP_BIND 44	51	FAD (ADP PART) (BY SIMILARITY).				

entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC IsoID=Q91740-1; Sequence=Displayed; TISSUE_SPECIFICITY: IN EARLY XENOPUS EMBRYO, CELLULAR FORMS OF FIBRONECTIN PREDOMINATE WHICH INCLUDE BOTH EXTRA DOMAINS, IN FIBRONECTIN OF BIONFORMATICS AND ADULT LIVER THE CONNECTING STRAND 3 CAN BE EITHER COMPLETELY EXCLUDED OR INCLUDED.

CC -!- SIMILARITY: Contains 12 fibronectin type I domains.

CC -!- SIMILARITY: Contains 2 fibronectin type II domains.

CC -!- SIMILARITY: Contains 17 fibronectin type III domains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC DR M77820; AR49707.1; -.

CC DR HSSP; P02751; FN2.

CC DR InterPro; IPR001356; Homeobox.

CC DR InterPro; IPR000912; Octamer-bind_TF.

CC DR InterPro; IPR000327; POU_domain.

CC DR InterPro; IPR007103; POU_homeo.

CC DR Pfam; PF000046; homeobox; 1.

CC DR Pfam; PF00157; pou; 1.

CC DR Prints; PR00028; PODOMAIN.

CC DR Prodrom; PD00010; Homeobox; 1.

CC DR SMART; SMD00389; HOX; 1.

CC DR SMART; SMD00352; POU; 1.

CC DR PROSITE; PS00027; HOMEROBOX_1; 1.

CC DR PROSITE; PS00035; POU_1; 1.

CC DR PROSITE; PS00045; POU_2; 1.

CC DR PROSITE; PS50071; HOMEROBOX_2; 1.

CC KW Homeobox; DNA-binding; Transcription regulation; Activator;

CC KW Nuclear protein.

FT DOMAIN 398 POU.

FT DOMAIN 397 456 HOMEobox.

SQ SEQUENCE 760 AA; 79097 MW; 7066369DA9018345 CRC64;

Query Match Best Local Similarity 71.4%; Score 35; DB 1; Length 760; Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 STAPPYHNV 9

Db 529 STAPPVSSV 537

RESULT 10

FINC_XENLA STANDARD; PRT; 2481 AA.

ID Q91740; STANDARD; PRT; 2481 AA.

AC 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Fibronectin precursor.

GN Xenopus laevis (African clawed frog).

OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Xenopoda; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;

OC Xenopodinae; Xenopus.

RN SEQUENCE FROM N.A. MEDLINE=92111942; PubMed=1730390;

RX Desimone D.W., Norton P.O., Hynes R.O.; "Identification and characterization of alternatively spliced fibronectin mRNAs expressed in early Xenopus embryos." ; Dev. Biol. 149:357-369(1992).

CC -!- FUNCTION: Fibronectins bind cell surfaces and various compounds including collagen, fibrin, heparin, DNA, and actin. Fibronectins are involved in cell adhesion, cell motility, opsonization, wound healing, and maintenance of cell shape (By similarity).

CC -!- CONNECTED BY DISULFIDE BONDS NEAR THE CARBOXYL ENDS (By SIMILARITY).

CC -!- ALTERNATIVE PRODUCTS:

CC Event=Alternative splicing; Named isoforms; Comment=A number of isoforms are produced. Each of the "extra domain" and the connecting strand 3 are present in some forms of Fibronectin and absent in others; Name=1;

FT DOMAIN	2299	2343	FIBRONECTIN TYPE-I 10.	RT Xenopus, that is related to AIB1, p/CIP and TIF2. ";
FT DOMAIN	2344	2386	FIBRONECTIN TYPE-I 11.	RL Mol. Endocrinol. 12:1103-1047(1998).
FT DOMAIN	2388	2431	FIBRONECTIN TYPE-I 12.	- - FUNCTION: Nuclear receptor coactivator that directly binds nuclear receptors and stimulates the transcriptional activities in a hormone-dependent fashion. Plays a central role in creating a multisubunit coactivator complex, probably via remodeling of chromatin. Involved in the coactivation of different nuclear receptors, such as retinoids (RAR and RXR), thyroid hormone (TR) and orphan nuclear receptor (hepatocyte nuclear receptor 4 (HNF4)).
FT SITE	1615	1617	CELL ATTACHMENT SITE.	CC
FT DISULFID	55	81	BY SIMILARITY.	CC
FT DISULFID	79	90	BY SIMILARITY.	CC
FT DISULFID	100	128	BY SIMILARITY.	CC
FT DISULFID	126	138	BY SIMILARITY.	CC
FT DISULFID	144	172	BY SIMILARITY.	CC
FT DISULFID	170	182	BY SIMILARITY.	CC
FT DISULFID	189	218	BY SIMILARITY.	CC
FT DISULFID	216	228	BY SIMILARITY.	CC
FT DISULFID	234	263	BY SIMILARITY.	CC
FT DISULFID	261	273	BY SIMILARITY.	CC
FT DISULFID	309	336	BY SIMILARITY.	CC
FT DISULFID	334	343	BY SIMILARITY.	CC
FT DISULFID	361	387	BY SIMILARITY.	CC
FT DISULFID	375	402	BY SIMILARITY.	CC
FT DISULFID	421	447	BY SIMILARITY.	CC
FT DISULFID	471	499	BY SIMILARITY.	CC
FT DISULFID	497	509	BY SIMILARITY.	CC
FT DISULFID	519	546	BY SIMILARITY.	CC
FT DISULFID	544	556	BY SIMILARITY.	CC
FT DISULFID	562	590	BY SIMILARITY.	CC
FT DISULFID	588	600	BY SIMILARITY.	CC
FT DISULFID	2301	2330	BY SIMILARITY.	CC
FT DISULFID	2328	2340	BY SIMILARITY.	CC
FT DISULFID	2346	2373	BY SIMILARITY.	CC
FT DISULFID	2371	2383	BY SIMILARITY.	CC
FT DISULFID	2390	2414	BY SIMILARITY.	CC
FT DISULFID	2412	2428	BY SIMILARITY.	CC
FT DISULFID	2459	2459	INTERCHAIN (WITH C-2463) (BY SIMILARITY).	CC
FT DISULFID	2463	2463	N-LINKED (GLCNAC. . .) (POTENTIAL).	CC
FT CARBOHD	431	431	N-LINKED (GLCNAC. . .) (POTENTIAL).	CC
FT CARBOHD	529	529	N-LINKED (GLCNAC. . .) (POTENTIAL).	CC
FT CARBOHD	543	543	N-LINKED (GLCNAC. . .) (POTENTIAL).	CC
FT CARBOHD	877	877	N-LINKED (GLCNAC. . .) (POTENTIAL).	CC
FT CARBOHD	1244	1244	N-LINKED (GLCNAC. . .) (POTENTIAL).	CC
FT CARBOHD	1291	1291	N-LINKED (GLCNAC. . .) (POTENTIAL).	CC
FT CARBOHD	2202	2202	N-LINKED (GLCNAC. . .) (POTENTIAL).	CC
SQ SEQUENCE	2481 AA:	272678 MW:	7E4DF46CE72C93 CRC64;	
				CC
Query Match		69.44%	Score 34; DB 1; Length 2481;	DR EMBL; AF04408; AAC1297.1; -.
Best Local Similarity		85.77%	Pred. No. 3.2e+02;	DR InterPro; IPR001092; HLH basic.
Matches	6;	Conservative	0; Mismatches 1; Indels 0; Gaps 0;	DR PROSITE; PSS0112; PAS; 1.
Qy	3 APPHEHV 9			DR KW Transferase; Transcription regulation; Activator; Nuclear protein; DR InterPro; IPR000014; PAS-domain.
Db	813 APPHEHV 819			DR Acetylation; Phosphorylation; Repeat; Alternative splicing.
RESULT 11				CC
NCI3_XENLA	STANDARD;	ERT;	1391 AA.	CC
AC 057539;	(Rel. 41, Created)			CC
DT 28-FEB-2003	(Rel. 41, Last sequence update)			CC
DT 28-FEB-2003	(Rel. 42, Last annotation update)			CC
DT 10-OCT-2003	(Rel. 42, Last annotation update)			CC
DE interacting coactivator xSRC-3.				CC
GN NCOA3.				CC
OS Xenopus laevis (African clawed frog)				CC
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				CC
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;				CC
OC Xenopoda; Xenidae; Xenopus.				CC
OX NCBI_TAXID=3555;				CC
RN [1]				CC
RP SEQUENCE FROM N.A.' INTERACTION WITH RXRA, THRA AND EP300, AND MUTAGENESIS OF LEU-622; LEU-683 AND LEU-739.				CC
TISSUE-OOCYTE:				CC
RC PMID=9832543; PubMed=9558407;				CC
RA Kim H.-J., Lee S.-K., Na S.-Y., Choi H.-S., Lee J.W.;				CC
RA "Molecular cloning of xSRC-3, a novel transcription coactivator from				CC
RA L->A: STRONGLY IMPAIRS INTERACTION WITH				CC

NUCLEAR RECEPTORS.
L->A: STRONGLY IMPAIRS INTERACTION WITH
NUCLEAR RECEPTORS.

PT MUTAGEN 739 739
FT SEQUENCE 1391 AA; 152532 MW; AD285CD934AC33D CRC64;
Query Match Score 33; DB 1; Length 1391;
Best Local Similarity 67.3%; Pred. No. 2.7e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy 1 STAPPVHN 9
Db 860 SSTPPVRN 868

RESULT 1.2
POLR_TVMVA STANDARD PRT; 1844 AA.
ID_POLR_TVMVA
DT_01-FEB-1991 (Rel. 1.7, Created)
DT_01-FEB-1991 (Rel. 1.7, Last sequence update)
DT_28-FEB-2013 (Rel. 4.1, Last annotation update)
DE RNA replicase polyprotein (EC 2.7.7.18).
OS Turnip yellow mosaic virus (Australian isolate).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tymoviridae;
OC Virus;
OX NCBI_TaxID=12155;
RN

SEQUENCE FROM N.A.
MEDLINE=90021184; PubMed=2800335;
RX
RA Keese P.; Mackenzie A.; Gibbs A.;
RT Nucleotide sequence of the genome of an Australian isolate of turnip
RT yellow mosaic tymovirus";
RL Virology 172:536-546 (1989).
CC -|- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
(RNA) (N)
CC -|- MISCELLANEOUS: THE 206 kDa PROTEIN IS POTENTIALLY A POLYPROTEIN
(BY HOMOLOGY TO THE LONGER PROTEIN OF THE SINDBAS VIRUS).
CC

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR PIR; J00109; BRWPTM.
DR InterPro; IPR008943; Peptidase_C21.
DR InterPro; IPR001788; RNA_dep_RNapol2.
DR InterPro; IPR007095; RNA_pol_DS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR InterPro; IPR006060; Viral_helicasel.
DR PFam; PF05381; Peptidase_C21_1.
DR PFam; PF00918; RNA_dep_RNapol2_1.
DR PFam; PF01443; Viral_helicasel_1.
KW Transferase; RNA directed RNA polymerase; Polyprotein; ATP-binding.
FT NE BIND 976 983 ATP (BY SIMILARITY).
FT DOMAIN 1631 1665 POLYMERASE SITE (BY SIMILARITY).
SQ SEQUENCE 1844 AA; 206510 MW; CB447EFO5F199A18_CRC64;

Query Match Score 33; DB 1; Length 1844;
Best Local Similarity 67.3%; Pred. No. 3.6e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 1 STAPPVHN 8
Db 581 STAPPDH5 588

RESULT 13
PR33_DAUC_A STANDARD PRT; 211 AA.

nucliosome chains into higher order structures.

-|- SUBCELLULAR LOCATION: Nuclear.

CC -|- SIMILARITY: Belongs to the histone H1/H5 Family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to licensee@isb-sib.ch).

DR EMBL; X17072; CAA34948.1; - .

DR EMBL; X52576; CAA36804.1; - .

DR PIR; S21935; S21935.

DR HSSP; P02259; IHS1.

DR FlyBase; Fgn0012374; Dhyd\His1.

DR InterPro; IPR005811; Histone_H1/H5.

DR InterPro; IPR005819; Histone_H5.

DR Pfam; PF00338; Linker histone; 1.

DR PRINTS; PRO0624; HISTONEHRS.

DR SMART; SM00526; HIS_1.

KW Chromosomal protein; Nuclear protein; DNA-binding.

FT CONFLICT 208 208 A -> P (IN REF. 2).

SEQUENCE 249 AA; 25746 MW; 189236C0518F903 CRC64;

Query Match 65.3% Score 32; DB 1; Length 249;

Best Local Similarity 71.4% Pred. No. 69; 1; Mismatches 1; Indels 0; Gaps 0; Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 STAPPVH 7

Db 40 TMAPPH 46

RESULT 15

CCO4_HUMAN STANDARD; PRT; 265 AA.

AC O9Y3AO_Q9CEWA4.

DT 28-FEB-2003 (Rel. 41, Created)

DT 10-OCT-2003 (Rel. 42, Last sequence update)

DE Ubiquinone biosynthesis protein COQ4 homolog (coenzyme Q biosynthesis protein 4 homolog) (CGI-92).

GN CCO4.

OS Homo sapiens (Human).

OC Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

CX NCBITaxonID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE:20227150; PubMed=10810093;

RA Lai C.-H., Chou C.-Y., Chang L.-Y., Liu C.-S., Lin W.-C.;

RT "Identification of novel human genes evolutionarily conserved in Caenorhabditis elegans by comparative proteomics.";

RR Genome Res. 10:703-713 (2000).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE:22388257; PubMed=12477932;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shemesh C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Blat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Dratchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Usdin T.B., Toshiyuki S., Carrascosa P., Orange C., Raha S.S., Loquelandano N.A., Peters G.J., Carrington R.D., Mullally S.J., Bozak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalon D.K., Manly D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahy J., Heitton B., Kettman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bonifard G.G., RA

GenCore version 5.1.6
 Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 28, 2004, 18:08:33 ; Search time 35 Seconds
 (without alignments)
 81.133 Million cell updates/sec

Title: US-10-019-513-1
 Perfect score: 49
 Sequence: 1 STAPPYHNV 9

Scoring table: BLOSUM62
 Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%
 Listing first 45 summaries

Database : SPTREMBL_25;*

1: sp_archea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mc:*

8: sp_organelle:*

9: sp_Phage:*

10: sp_Plant:*

11: sp_rrodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_virus:*

16: sp_bacteria:*

17: sp_archeap:*

17 36 73.5 12 Q8v7e1 tt virus. o
 18 36 73.5 12 Q8vrg2 tt virus. o
 19 36 73.5 12 Q9dih9 tt virus. o
 20 36 73.5 12 Q9gax4 tt virus. a
 21 36 73.5 12 Q9duo tt virus. o
 22 36 73.5 12 Q9qa81 tt virus. o
 23 36 73.5 12 Q9wsx8 tt virus. o
 24 36 73.5 12 Q7vcs5 tt virus. o
 25 36 73.5 12 Q8v7f5 tt virus. o
 26 36 73.5 12 Q8v7f6 tt virus. o
 27 36 73.5 305 12 Q8qnq4 ectocarpus
 28 36 73.5 329 10 Q9st6 arabidopsis
 29 36 73.5 335 16 Q9Exq5 mycobacteris
 30 36 73.5 341 10 Q9frk5
 31 36 73.5 391 16 Q7u9n5
 32 36 73.5 531 5 Q7r8q4
 33 36 73.5 690 3 Q06993
 34 36 73.5 716 13 Q7zvq5
 35 36 73.5 720 5 Q3847
 36 36 73.5 1013 5 Q9677s
 37 36 73.5 4001 5 Q8wro7
 38 36 73.5 4001 5 Q9vcab8
 39 36 73.5 4793 3 Q7zbp4
 40 35 71.4 128 12 Q8v7e8 tt virus. o
 41 35 71.4 144 11 Q7ct08 mus musculu
 42 35 71.4 157 16 Q8t2s8 streptomyce
 43 35 71.4 182 5 Q19136 caenorhabdi
 44 35 71.4 235 10 Q9ffy9 zea mays (m
 45 35 71.4 253 16 Q8ib73 bacillus ce

ALIGNMENTS

RESULT 1

Q8YX80 PRELIMINARY; PRT; 590 AA.
 ID Q8YX80; AC Q8YX80;
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
 DE Hypothetical protein All8037.
 GN All8037.
 OS Anabaena sp. (strain PCC 7120).
 OG Plasmid pCC7120gamma.
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
 NCBI_TaxID:103690; [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE-2159525; PubMed=11759840;
 RX KANEKO T., NAKAMURA Y., WOLK C.P., KURITZ T., SASAMOTO S.,
 RA WATANABE A., IRIGUCHI M., ISHIKAWA A., KAWASHIMA K., KIMURA T.,
 RA KISHIDA Y., KONARA M., MATSUMOTO M., MATSUO A., MURAKI A.,
 RA NAKAZAKI N., SHIMPO S., SUGIMOTO M., TAKAZAWA M., YAMADA M.,
 RA YASUDA M., TABATA S.
 RT "Complete genomic sequence of the filamentous nitrogen-fixing
 cyanobacterium Anabaena sp. strain PCC 7120.";
 RT DNA RES. B:205-213 (2001);
 DR EMBL; AP003603; BAB77367.1; -.
 DR PIR; AF25553; AF25537.
 DR GO; GO:0046321; C:extrachromosomal DNA; IEA.
 DR GO; GO:0016620; C:membrane; IEA.
 DR GO; GO:0009391; P:unidirectional conjugation; IEA.
 DR InterPro; IPR003688; TRAG.
 DR Pfam; PF02534; TRAG; 1.
 KW Plasmid; Hypothetical protein; Complete proteome.
 SQ SEQUENCE 590 AA; 64933 MW; 7EC468DF65E3B459 CRC64;

Query Match 83.7%; Score 41; DB 16; Length 590;
 Best Local Similarity 87.5%; Pred. No. 10;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Q9j974 tt virus.

Db	9 TAPPAHNV 16		Query Match Best Local Similarity 81.6%; Matches 7; Conservative 0;	Score 40; DB 5; Length 803;	PRT; 803 AA;	88900 MW;	3F0752DCE4B32CA4 CRC64;
RESULT 2							
Q9MZL1	PRELIMINARY;	PRT; 553 AA.	Q9UMI8	PRELIMINARY;	PRT;	20 AA.	
AC Q9MZL1			ID Q9UMI8				
DT 01-OCT-2000	(TREMBLrel. 15, Created)		AC Q9UMI8;				
DT 01-OCT-2000	(TREMBLrel. 15, Last sequence update)		DT 01-MAY-2000 (TREMBLrel. 13, Created)				
DT 01-MAR-2003	(TREMBLrel. 23, Last annotation update)		DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)				
DE Mucin 1 (Fragment).			DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)				
OS Macaca mulatta (Rhesus macaque).			DE Mucin (Fragment).				
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			OS Homo sapiens (Human).				
OC Cercopithecoidea; Primates; Catarrhini; Hominoidea; Homo.			OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC Cercopithecinae; Macaca.			OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
ON NCBI_TaxID=9544;			RN [1]				
RN [1]			RP SEQUENCE FROM N.A.				
RP SEQUENCE FROM N.A.			RX MEDLINE=89235154; PubMed=2715633;				
RX MEDLINE=20330533; PubMed=10869775;			RA Xing, P.X.; Tijandria, J.J.; Reynolds, K.; McLaughlin, P.J.; Purcell, D.F.J.,				
RX Pieterz G.A.;			RA McKenzie, I.P.C.;				
RT Mannan"; ";			RT "Reactivity of anti-human milk fat globule antibodies with synthetic				
RT Mannan"; ";			RT peptides.";				
RT Mannan"; ";			RL J. Immunol. 142:15003-1509 (1989).				
RL Vaccine 18:3297-3309 (2000);			DR EMBL; M26316; AAA36336.1; -.				
DR EMBL; AF176347; ZAF8243.1;			DR PTR; S10571; S10571.				
DR InterPro; IPR00104; Crystallin.			FT NON-TER 1 1 1				
DR InterPro; IPR00002; SEA_domain.			FT NON-TER 20 20 20				
DR Pfam; PF01390; SEA; 1.			SQ SEQUENCE 20 AA; 1887 MW;				
DR SMART; SM00220; SEA; 1.			5B3473BAEBAFAD87 CRC64;				
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 1.			Query Match Best Local Similarity 79.6%; Matches 7; Conservative 77.8%;				
DR PROSITE; PS0024; SEA; 1.			Score 39; DB 4; Length 20;				
FT NON-TER 1 1			Pred. No. 0.74%;				
FT NON-TER 553 AA; 55778 MW;			Mismatches 2; Indels 0; Gaps 0;				
SQ SEQUENCE 553 AA; 55778 MW;							
RESULT 3							
Q7YSX2	PRELIMINARY;	PRT; 803 AA.	Q9SFH0	PRELIMINARY;	PRT;	499 AA.	
AC Q7YSX2;			ID Q9SFH0				
DT 01-OCT-2003	(TREMBLrel. 25, Created)		AC Q9SFH0;				
DT 01-OCT-2003	(TREMBLrel. 25, Last sequence update)		DT 01-MAY-2000 (TREMBLrel. 13, Created)				
DT 01-OCT-2003	(TREMBLrel. 25, Last annotation update)		DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)				
DE Enoyl-CoA hydratase/enoyl-CoA isomerase/3-hydroxyacyl-CoA dehydrogenase, putative.			DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)				
GN TB97.2 .4130.			DE Gamma-glutamylcysteine synthetase precursor.				
OS Trypanosoma brucei.			GN GSHL.				
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatida; Trypanosoma.			OS Pisum sativum (Garden pea).				
NCBI_TaxID=5691;			OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
RN [1]			OC Spermatophyta; Magnoliophyta; euicots; core euicots; Rosids;				
RP SEQUENCE FROM N.A.			OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciaceae; Pisum.				
RC STRAIN=Gutta10.1.			RN [1]				
RA El-Sayed N.M.A.;	Ghedin E., Song J., MacLeod A., Bringaud F., Twedie A.,		RP SEQUENCE FROM N.A.				
RA Larkin C., Wanless D., Peterson J., Hou L., Taylor S., Twedie A.,			RC TISSUE=Nodule;				
RA Bitcau N., Khalak H.G., Lin X., Mason T., Hannick L., Caler E.,			RC Q9SHH0;				
RA Blandin G., Bartholomeu D., Simpson A.J., Kaul S., Zhao H., Pai G.,			AC Q9SHH0;				
RA Van Aken S., Utterback T., Haas B., Koo H.-L., Ohnayak L., Suh B.,			DT 01-MAY-2000 (TREMBLrel. 13, Created)				
RA Gerrard C., Leech V., Qi R., Zhou S., Schwartz D., Feldblum T.,			DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)				
RA Salzberg S., Tait A., Turner M.R., White O., Melville S.,			DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)				
RA Adams M.D., Fraser C.M., Doneelson J.E.;			DE Gamma-glutamylcysteine synthetase precursor.				
RA "The sequence and analysis of Trypanosoma brucei chromosome II."			GN GSHL.				
RI Nucleic Acids Res. 0:0-0(2003) .			OS Pisum sativum (Garden pea).				
RI EMBL: AE017169; AAQ15870.1; -.			OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
KW Isomerase.			OC Spermatophyta; Magnoliophyta; euicots; core euicots; Rosids;				
KW Transit peptide.			OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciaceae; Pisum.				
FT TRANSIT 1			RN [1]				
RESULT 4							
Q9UMI8	PRELIMINARY;	PRT;	Q9UMI8	PRELIMINARY;	PRT;	20 AA.	
AC Q9UMI8;			ID Q9UMI8				
DT 01-MAY-2000	(TREMBLrel. 13, Created)		AC Q9UMI8;				
DT 01-JUN-2003	(TREMBLrel. 24, Last annotation update)		DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)				
DE Mucin (Fragment).			DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)				
OS Homo sapiens (Human).			DE Gamma-glutamylcysteine synthetase precursor.				
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			GN GSHL.				
RN [1]			OS Pisum sativum (Garden pea).				
RP SEQUENCE FROM N.A.			OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
RC TISSUE=Nodule;			OC Spermatophyta; Magnoliophyta; euicots; core euicots; Rosids;				
RC Q9SHH0;			OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciaceae; Pisum.				
AC Q9SHH0;			RN [1]				
DT 01-MAY-2000	(TREMBLrel. 13, Created)		RP SEQUENCE FROM N.A.				
DT 01-JUN-2003	(TREMBLrel. 24, Last annotation update)		RC TISSUE=Nodule;				
DE Mucin (Fragment).			RC Q9SHH0;				
OS Homo sapiens (Human).			AC Q9SHH0;				
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			DT 01-MAY-2000 (TREMBLrel. 13, Created)				
RN [1]			DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)				
RP SEQUENCE FROM N.A.			DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)				
RC TISSUE=Nodule;			DE Gamma-glutamylcysteine synthetase precursor.				
RC Q9SHH0;			GN GSHL.				
AC Q9SHH0;			OS Pisum sativum (Garden pea).				
DT 01-MAY-2000	(TREMBLrel. 13, Created)		OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
DT 01-JUN-2003	(TREMBLrel. 24, Last annotation update)		OC Spermatophyta; Magnoliophyta; euicots; core euicots; Rosids;				
DE Mucin (Fragment).			OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciaceae; Pisum.				
OS Homo sapiens (Human).			RN [1]				
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			RP SEQUENCE FROM N.A.				
RN [1]			RC TISSUE=Nodule;				
RP SEQUENCE FROM N.A.			RC Q9SHH0;				
RC TISSUE=Nodule;			AC Q9SHH0;				
RC Q9SHH0;			DT 01-MAY-2000 (TREMBLrel. 13, Created)				
AC Q9SHH0;			DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)				
DT 01-MAY-2000	(TREMBLrel. 13, Created)		DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)				
DT 01-JUN-2003	(TREMBLrel. 24, Last annotation update)		DE Gamma-glutamylcysteine synthetase precursor.				
DE Mucin (Fragment).			GN GSHL.				
OS Homo sapiens (Human).			OS Pisum sativum (Garden pea).				
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
RN [1]			OC Spermatophyta; Magnoliophyta; euicots; core euicots; Rosids;				
RP SEQUENCE FROM N.A.			OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciaceae; Pisum.				
RC TISSUE=Nodule;			RN [1]				
RC Q9SHH0;			RP SEQUENCE FROM N.A.				
AC Q9SHH0;			RC TISSUE=Nodule;				
DT 01-MAY-2000	(TREMBLrel. 13, Created)		RC Q9SHH0;				
DT 01-JUN-2003	(TREMBLrel. 24, Last annotation update)		AC Q9SHH0;				
DE Mucin (Fragment).			DT 01-MAY-2000 (TREMBLrel. 13, Created)				
OS Homo sapiens (Human).			DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)				
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)				
RN [1]			DE Gamma-glutamylcysteine synthetase precursor.				
RP SEQUENCE FROM N.A.			GN GSHL.				
RC TISSUE=Nodule;			OS Pisum sativum (Garden pea).				
RC Q9SHH0;			OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
AC Q9SHH0;			OC Spermatophyta; Magnoliophyta; euicots; core euicots; Rosids;				
DT 01-MAY-2000	(TREMBLrel. 13, Created)		OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciaceae; Pisum.				
DT 01-JUN-2003	(TREMBLrel. 24, Last annotation update)		RN [1]				
DE Mucin (Fragment).			RP SEQUENCE FROM N.A.				
OS Homo sapiens (Human).			RC TISSUE=Nodule;				
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			RC Q9SHH0;				
RN [1]			AC Q9SHH0;				
RP SEQUENCE FROM N.A.			DT 01-MAY-2000 (TREMBLrel. 13, Created)				
RC TISSUE=Nodule;			DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)				
RC Q9SHH0;			DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)				
AC Q9SHH0;			DE Gamma-glutamylcysteine synthetase precursor.				
DT 01-MAY-2000	(TREMBLrel. 13, Created)		GN GSHL.				
DT 01-JUN-2003	(TREMBLrel. 24, Last annotation update)		OS Pisum sativum (Garden pea).				
DE Mucin (Fragment).			OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OS Homo sapiens (Human).			OC Spermatophyta; Magnoliophyta; euicots; core euicots; Rosids;				
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciaceae; Pisum.				
RN [1]			RN [1]				
RP SEQUENCE FROM N.A.			RP SEQUENCE FROM N.A.				
RC TISSUE=Nodule;			RC TISSUE=Nodule;				
RC Q9SHH0;			RC Q9SHH0;				
AC Q9SHH0;			AC Q9SHH0;				
DT 01-MAY-2000	(TREMBLrel. 13, Created)		DT 01-MAY-2000 (TREMBLrel. 13, Created)				
DT 01-JUN-2003	(TREMBLrel. 24, Last annotation update)		DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)				
DE Mucin (Fragment).			DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)				
OS Homo sapiens (Human).			DE Gamma-glutamylcysteine synthetase precursor.				
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			GN GSHL.				
RN [1]			OS Pisum sativum (Garden pea).				
RP SEQUENCE FROM N.A.			OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
RC TISSUE=Nodule;			OC Spermatophyta; Magnoliophyta; euicots; core euicots; Rosids;				
RC Q9SHH0;			OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciaceae; Pisum.				
AC Q9SHH0;			RN [1]				
DT 01-MAY-2000	(TREMBLrel. 13, Created)		RP SEQUENCE FROM N.A.				
DT 01-JUN-2003	(TREMBLrel. 24, Last annotation update)		RC TISSUE=Nodule;				
DE Mucin (Fragment).			RC Q9SHH0;				
OS Homo sapiens (Human).			AC Q9SHH0;				
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			DT 01-MAY-2000 (TREMBLrel. 13, Created)				
RN [1]			DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)				
RP SEQUENCE FROM N.A.			DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)				
RC TISSUE=Nodule;			DE Gamma-glutamylcysteine synthetase precursor.				
RC Q9SHH0;			GN GSHL.				
AC Q9SHH0;			OS Pisum sativum (Garden pea).				
DT 01-MAY-2000	(TREMBLrel. 13, Created)		OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
DT 01-JUN-2003	(TREMBLrel. 24, Last annotation update)		OC Spermatophyta; Magnoliophyta; euicots; core euicots; Rosids;				
DE Mucin (Fragment).			OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciaceae; Pisum.				
OS Homo sapiens (Human).			RN [1]				
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			RP SEQUENCE FROM N.A.				
RN [1]			RC TISSUE=Nodule;				
RP SEQUENCE FROM N.A.			RC Q9SHH0;				
RC TISSUE=Nodule;			AC Q9SHH0;				
RC Q9SHH0;			DT 01-MAY-2000 (TREMBLrel. 13, Created)				
AC Q9SHH0;			DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)				
DT 01-MAY-2000	(TREMBLrel. 13, Created)		DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)				
DT 01-JUN-2003	(TREMBLrel. 24, Last annotation update)		DE Gamma-glutamylcysteine synthetase precursor.				
DE Mucin (Fragment).			GN GSHL.	</td			

Qy	1	STAPPVHN 8	79.6%; Score: 39; DB 13; Length: 802;	Indels: 0; Gaps: 0;	RN [1]_RN [1]	RP
Db	10	STAPPVHN 17				
RESULT 6						
O42127		PRELIMINARY;	PRT; 802 AA.			
AC	O42127	01-JAN-1998 (TREMBLrel. 05, Created)				
DT	01-OCT-2003 (TREMBLrel. 25, Last annotation update)					
DB	EGF Receptor 3.					
OS	Xenopus laevis (African clawed frog).					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;					
OC	Xenopoda; Xenopus.					
RN	[1]_RN [1]	SEQUENCE FROM N.A.				
RP						
RA	Hongo I., Kengaku M., Okamoto H..	"Differential employment of FGF signaling system for the embryonic induction.", Submitted (SEP-1997) to the ENSEMBL/GenBank/DDBJ databases.				
RT						
RT						
RL	EMBL; AB007035; BAA2281.1; -.					
DR	HSSP; P13622; 1FGK.					
DR	GO; GO:0005524; F:ATP binding; IEA.					
DR	GO; GO:0004713; F:protein-tyrosine kinase activity; IEA.					
DR	GO; GO:0004872; F:receptor activity; IEA.					
DR	GO; GO:0016740; F:transferrase activity; IEA.					
DR	GO; GO:0006468; P:protein amino acid phosphorylation; IEA.					
DR	InterPro; IPR007110; Ig-like.					
DR	InterPro; IPR003598; Ig_C2.					
DR	InterPro; IPR000719; Prot_kinase.					
DR	InterPro; IPR001245; Tyr_pkirase.					
DR	InterPro; IPR008265; Tyr_pkirase_AS.					
PFam	PF00047; ig_1.					
DR	PRNTS; ER00109; pkirase_1.					
DR	PRODom; ED000001; Prot_kinase.					
SMART	SM00219; IgGc2_2.					
DR	PROSITE; PS50835; Ig_LIKE; 3.					
DR	PROSITE; PS500107; PROTEIN_KINASE_ATP; 1.					
DR	PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.					
DR	PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.					
KW	ATP-binding; Immunoglobulin domain; Kinase; Receptor; Transferase; Tyrosine-protein kinase.					
SQ	SEQUENCE 802 AA; 89515 MW; CC5E5DDFB3D25BD3 CRC64;					
Query Match	79.6%; Score: 39; DB 13; Length: 802;					
Best Local Similarity	87.5%; Pred. No. 36;					
Matches	7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;					
Qy	2 TAPPVHN 9					
Db	396 TAPPVHKV 403					
RESULT 7						
Q9VC000		PRELIMINARY;	PRT; 2768 AA.			
ID	Q9VC00					
AC	Q9VC00;					
DT	01-MAY-2000 (TREMBLrel. 13, Created)					
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)					
DT	01-JUN-2003 (TREMBLrel. 24, Last annotation update)					
DE	CG13648 protein.					
Query Match	79.6%; Score: 39; DB 5; Length: 2768;					
Best Local Similarity	77.8%; Pred. No. 1.e+02;					
Matches	7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;					
Qy	1 STAPPVHN 9					
Db	2332 STAPPVHDY 2340					
RESULT 8						
Q14306		PRELIMINARY;	PRT; 1191 AA.			
ID	Q14306					
AC	Q14306;					
DT	01-JUN-1998 (TREMBLrel. 06, Created)					
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)					
DT	01-OCT-2003 (TREMBLrel. 25, Last annotation update)					
Query Match	79.6%; Score: 39; DB 5; Length: 2768;					
Best Local Similarity	77.8%; Pred. No. 1.e+02;					
Matches	7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;					
Qy	1 STAPPVHN 9					
Db	2332 STAPPVHDY 2340					

DE	Putative inositol polyphosphate phosphatase.	DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)	
GN	SPAC9G1.10C.	DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)	
OS	Schizosaccharomyces pombe (Fission yeast).	DE	ORF2 protein (Fragment).	
OC	Bukaryota; Fungi; Ascomycota; Schizosaccharomycetes;	OS	TT virus.	
OC	Schizosaccharomycetales; Schizosaccharomycetaceae;	OC	Viruses; ssDNA viruses; Circoviridae; Anellovirus.	
OC	Schizosaccharomycetes.	OX	NCBI_TaxID=68887;	
OX		RN	SEQUENCE FROM N.A.	
RN		RP	SEQUENCE FROM N.A.	
RC	STRAIN=C972h;	RC	STRAIN=TREMBLrel. 12;	
RA	Churcher C.M., Gentles S., Barrell B.G., Rajandream M.A., Wood V.;	RA	MEDLINE=20129505; PubMed=10664406;	
RA	Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.	RA	Mulyantno, Hijikata M., Matsushita M., Ingkukusmo G., Widjaya A., Sumarsidi D., Kanai K., Ohta Y., Mishiro S.;	
RL	EMBL; 298763; CAB31494.2; -.	RA	"TT virus (TtV) genotypes in native and non-native prostitutes of Irian Jaya, Indonesia: implication for non-occupational transmission." Arch. Virol. 145:63-72 (2000).	
DR	FIR; T39233; T39233.	DR	Arch. Virol. 145:63-72 (2000).	
DR	GenedB SPOMB; SBAC9G1.10C; -.	DR	DR InterPro; IPR004437; BAA8610.1; -.	
DR	GO: GO:0004437; P:inositol_phosphatidylinositol_phosphatase_a. . . ; IEA.	DR	DR InterPro; IPR004437; TT_ORF2; 1.	
DR	InterPro; IPR005135; BxO_endo_phos.	DR	Pfam; PF02957; TT_ORF2; 1.	
DR	Pfam; PF03372; BxO_endo_phos; 1.	FT	NON_TER 108 108	
DR	SMART; SM00128; IPPC; 1_phos; 1.	SQ	SEQUENCE 108 AA; 12062 MN; D0C3E628DE29D3F7 CRC64;	
SQ	SEQUENCE 1191 AA; 131274 MW; 7BA86AD85F3AEB7 CRC64;	Query	Match 73.5%; Score 36; DB 12; Length 108;	
Query Match 75.5%; Score 37; DB 3; Length 1191;	Best Local Similarity 100.0%; Pred. No. 17;	Best Local Similarity 100.0%; Pred. No. 17;		
Best Local Similarity 77.8%; Pred. No. 1.4e+02;	Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;	QY	4 PPVHN 9		
QY	1 STAPPVHN 9	Db	33 PPVHN 38	
Db	253 STAPPVHN 261			
RESULT 9		RESULT 11		
Q9QIWO	SEQUENCE FROM N.A.	Q9WAY6	SEQUENCE FROM N.A.	
ID	PRELIMINARY;	ID	PRELIMINARY;	
Q9QIWO	PRT;	Q9WAY6	PRT;	
AC	96 AA.	AC	115 AA.	
DT	09QIWO; (TREMBLrel. 13, Created)	DT	Q9WAY6; (TREMBLrel. 12, Created)	
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)	DT	01-NOV-1999 (TREMBLrel. 12, Last sequence update)	
DT	01-MAY-2000 (TREMBLrel. 19, Last annotation update)	DT	01-NOV-1999 (TREMBLrel. 12, Last annotation update)	
DE	ORF2 protein (Fragment).	DE	01-DEC-2001 (TREMBLrel. 19, Last annotation update)	
DE	TT virus.	DE	ORF2 (Fragment).	
OS	Viruses; ssDNA viruses; Circoviridae; Anellovirus.	OS	TT virus.	
OC	NCBI_TaxID=68887;	OC	Viruses; ssDNA viruses; Circoviridae; Anellovirus.	
OX		OX	NCBI_TaxID=68887;	
RN		RN	[1].	
RP	SEQUENCE FROM N.A.	RP	SEQUENCE FROM N.A.	
RC	STRAIN=TT_WAM93;	RC	STRAIN=OTNAk-2-4G;	
RA	MEDLINE=0129505; PubMed=10664406;	RA	DR InterPro; IPR004437; BAA7744.1; -.	
RA	Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.	RA	DR InterPro; IPR004437; TT_ORF2.	
RA	EMBL; AB01733; AB04118; TT_ORF2.	RA	DR Pfam; PF02957; TT_ORF2; 1.	
RA	InterPro; IPR004118; TT_ORF2.	FT	NON_TER 115 115	
RA	Pfam; PF02957; TT_ORF2; 1.	SQ	SEQUENCE 115 AA; 12324 MN; E4B2702604E41109 CRC64;	
FT	NON_TER 1 1	Query	Match 73.5%; Score 36; DB 12; Length 115;	
SQ	SEQUENCE 96 96 AA; 11160 MW; 12FF57A821FB2871 CRC64;	Best Local Similarity 100.0%; Pred. No. 18;	Best Local Similarity 100.0%; Pred. No. 18;	
Query Match 73.5%; Score 36; DB 12; Length 96;	Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
Best Local Similarity 100.0%; Pred. No. 15;	QY	4 PPVHN 9	Q8V7B8 PRELIMINARY; PRT; 127 AA.	
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	Db	36 PPVHN 41	Db	33 PPVHN 38
RESULT 12		RESULT 12		
Q9QTY6	SEQUENCE FROM N.A.	Q9QTY6	SEQUENCE FROM N.A.	
ID	PRELIMINARY;	ID	PRELIMINARY;	
Q9QTY6	PRT;	Q9QTY6	PRT;	
AC	108 AA.	AC	108 AA.	
DT	01-MAR-2002 (TREMBLrel. 20, Created)	DT	01-MAR-2002 (TREMBLrel. 20, Last sequence update)	
DT	01-MAR-2002 (TREMBLrel. 20, Last annotation update)	DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)	

DE SEQUENCE FROM N.A.
 OS TT virus; ssDNA viruses; Circoviridae; Anellovirus.
 OX NCBI_TaxID=68887;
 CO
 RN SEQUENCE FROM N.A.
 RX MEDLINE=21844401; PubMed=11855633;
 RA Peng Y.H.; Nishizawa T.; Takahashi M.; Ishikawa T.; Yoshiikawa A.,
 RA Okamoto H.;
 RT "Analysis of the complete genomes of thirteen TT virus variants
 classifiable into the fourth and fifth genetic groups, isolated from
 viremic infants.";
 RT Arch. Virol. 147:21-41 (2002).
 RL EMBL; AB064610; BAB79363.1;
 RA
 RT
 DR InterPro; IPR004118; TT_ORF2.
 PFam; PF02957; TT_ORF2_1.
 DR
 SQ SEQUENCE 127 AA; 13395 MW; 66R858DAS65B625D CRC64;
 Qy Query Match 73.5%; Score 36; DB 12; Length 128;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db
 Qy 4 PPVHN 9
 Db 5 PPVHN 10

RESULT 15
 Q8V7D5 PRELIMINARY; PRT; 128 AA.
 ID Q8V7D5
 AC Q8V7D5;
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE ORF2.
 DS TT virus.
 CC Viruses; ssDNA viruses; Circoviridae; Anellovirus.
 CO NCBI_TaxID=68887;
 RN [1] -
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21844401; PubMed=11855633;
 RA Peng Y.H.; Nishizawa T.; Takahashi M.; Ishikawa T.; Yoshiikawa A.,
 RA Okamoto H.;
 RT "Analysis of the complete genomes of thirteen TT virus variants
 classifiable into the fourth and fifth genetic groups, isolated from
 viremic infants.";
 RT Arch. Virol. 147:21-41 (2002).
 RL EMBL; AB064610; BAB79363.1;
 RA
 RT
 DR InterPro; IPR004118; TT_ORF2.
 PFam; PF02957; TT_ORF2_1.
 DR
 SQ SEQUENCE 128 AA; 12971 MW; 0759F8CF3F7D0393 CRC64;
 Qy Query Match 73.5%; Score 36; DB 12; Length 128;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db
 Qy 4 PPVHN 9
 Db 5 PPVHN 10

Search completed: July 28, 2004, 18:10:40
 Job time : 35 secs

RESULT 14
 Q8V7F1 PRELIMINARY; PRT; 128 AA.
 ID Q8V7F1
 AC Q8V7F1;
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DE ORF2.
 DS TT virus.
 CO NCBI_TaxID=68887;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 28, 2004, 18:09:33 ; Search time 19 Seconds
(Without alignments)
24.454 Million cell updates/sec

Title: US-10-019-513-1

Perfect score: 49

Sequence: 1 STAPPVHN 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued Patents Am.*
1: /cgn2_6_ptodata/2/iaa/SA COMB.pep.*
2: /cgn2_6_ptodata/2/iaa/SB COMB.pep.*
3: /cgn2_6_ptodata/2/iaa/SC COMB.pep.*
4: /cgn2_6_ptodata/2/iaa/SB COMB.pep.*
5: /cgn2_6_ptodata/2/iaa/BCTUS COMB.pep.*
6: /cgn2_6_ptodata/2/iaa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	49	100.0	1867	US-08-479-537A-5
2	49	100.0	1867	Sequence 5, Appli
3	49	100.0	1867	Sequence 5, Appli
4	49	100.0	2035	Sequence 5, Appli
5	49	100.0	2035	Sequence 2, Appli
6	49	100.0	2035	Sequence 2, Appli
7	79.6	9	1	US-08-479-537A-5
8	39	79.6	9	Sequence 55, Appli
9	39	79.6	9	Sequence 19, Appli
10	39	79.6	9	Sequence 45, Appli
11	39	79.6	16	Sequence 1, Appli
12	39	79.6	19	Sequence 19, Appli
13	39	79.6	19	Sequence 3, Appli
14	39	79.6	20	Sequence 9, Appli
15	39	79.6	20	Sequence 1, Appli
16	39	79.6	20	Sequence 32, Appli
17	39	79.6	20	Sequence 20, Appli
18	39	79.6	20	Sequence 1, Appli
19	39	79.6	20	Sequence 1, Appli
20	39	79.6	20	Sequence 3, Appli
21	39	79.6	20	Sequence 1, Appli
22	39	79.6	20	Sequence 34, Appli
23	39	79.6	20	Sequence 20, Appli
24	39	79.6	20	Sequence 1, Appli
25	39	79.6	20	Sequence 40, Appli
26	39	79.6	20	Sequence 9, Appli
27	39	79.6	20	Sequence 11, Appli

ALIGNMENTS

RESULT 1
US-08-479-537A-5
; Sequence 5, Application US/08479537A
; Patent No. 5861381
; GENERAL INFORMATION:
; APPLICANT: CHAMBON, Pierre
; APPLICANT: KIENY, Marie-Paule
; APPLICANT: LATHE, Richard
; APPLICANT: HARDEVENI, Mara
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/479,537A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 90/13101
; FILING DATE: 23-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/FR91/00835
; FILING DATE: 23-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/403,576
; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Teskin, Robin L.
; REGISTRATION NUMBER: 35,030
; REFERENCE/DOCKET NUMBER: 017753-025
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1867 amino acids

NAME/KEY: Peptide
 LOCATION: 128..1899
 OTHER INFORMATION: /note= "The amino acids spanning 128 to 1899 constitute a repeated region wherein the repeat 20 amino acids, 17 of which are fixed. The number of such repeats varies from 1 to 40."
 FEATURE:
 NAME/KEY: Peptide
 LOCATION: 134
 OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa Xaa which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA, or CGG; and Ala = GCT, GCC, GCA, or GCG."
 FEATURE:
 NAME/KEY: Peptide
 LOCATION: 144
 OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACP or ACG; and Asn = AAT or AAC."
 FEATURE:
 NAME/KEY: Peptide
 LOCATION: 147
 OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA, or CGG; and Ala = GCT, GCC, GCA, or GCG."
 OTHER INFORMATION:
 FEATURE:
 NAME/KEY: Peptide
 LOCATION: 1..21
 OTHER INFORMATION: /note= "Amino acids 1 to 21 are a 21 amino acid precursor sequence."
 OTHER INFORMATION:
 US-08-479-537A-2

Query 1 STAPPVHN 9
 Db 1730 STAPPVHN 1738

RESULT 5
 US-09-083-116-2
 ; Sequence 2, Application US/09083116
 ; Patent No. 6203795
 ; GENERAL INFORMATION:
 ; APPLICANT: CHAMON, Pierre
 ; APPLICANT: LATHE, Richard
 ; APPLICANT: HAREVENTI, Mara
 ; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
 ; NUMBER OF SEQUENCES: 5
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
 ; STREET: P. O. Box 1404
 ; CITY: Alexandria
 ; STATE: Virginia
 ; COUNTRY: United States
 ; ZIP: 22313-1404
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/083,116
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/479,537
 ; FILING DATE:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: WO PCT/FR91/00835

FILING DATE: 23-OCT-1991
 PRIORITY APPLICATION DATA: US 08/039,320
 PRIORITY APPLICATION NUMBER: US 08/039,320
 FILING DATE: 04-APR-1993
 PRIORITY APPLICATION DATA: US 08/403,576
 PRIORITY APPLICATION NUMBER: US 08/403,576
 FILING DATE: 14-MAR-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Taskin, Robin L.
 REGISTRATION NUMBER: 35,030
 REFERENCE/DOCKET NUMBER: 017753-025
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 836-6620
 TELEFAX: (703) 836-2021
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2035 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: Peptide
 FEATURE:
 NAME/KEY: Peptide
 LOCATION: 128..1899 /note= "The amino acids spanning 128 to 1899 constitute a repeated region wherein the repeat 20 amino acids, 17 of which are fixed. The number of such repeats varies from 1 to 40."
 FEATURE:
 NAME/KEY: Peptide
 LOCATION: 134
 OTHER INFORMATION: /note= "Amino acid 134 is Xaa which is the codon for Pro or Ala wherein Pro = ACT, ACC, ACP or ACG; and Asn = AAT or AAC."
 FEATURE:
 NAME/KEY: Peptide
 LOCATION: 144
 OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa which is the codon for Thr or Asn wherein Thr = GCT, GCC, GCA, or GCG."
 OTHER INFORMATION:
 FEATURE:
 NAME/KEY: Peptide
 LOCATION: 1..21
 OTHER INFORMATION: /note= "Amino acids 1 to 21 are a 21 amino acid precursor sequence."
 OTHER INFORMATION:
 US-09-083-116-2

Query Match Score 49; DB 2; Length 2035;
 Best Local Similarity 100.0%; Pred. No. 4.4%;
 Matches 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 STAPPVHN 9
 Db 1730 STAPPVHN 1738

RESULT 6
 US-09-134-916A-2
 ; Sequence 2, Application US/09134916A
 ; Patent No. 632895
 ; GENERAL INFORMATION:
 ; APPLICANT: CHAMON, Pierre
 ; APPLICANT: LATHE, Richard
 ; APPLICANT: HAREVENTI, Mara
 ; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
 ; NUMBER OF SEQUENCES: 5
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
 ; STREET: P. O. Box 1404
 ; CITY: Alexandria
 ; STATE: Virginia
 ; COUNTRY: United States
 ; ZIP: 22313-1404
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/083,116
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/479,537
 ; FILING DATE:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: WO PCT/FR91/00835

NUMBER OF SEQUENCES: 5
 CORRESPONDENCE ADDRESS:
 STREET: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
 CITY: Alexandria
 STATE: Virginia
 COUNTRY: United States
 ZIP: 22313-1404
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/134,916A
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/08/479,537
 FILING DATE: 07-JUN-1995
 APPLICATION NUMBER: FR 90/13101
 FILING DATE: 23-OCT-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: WO PCT/FR91/00835
 FILING DATE: 23-OCT-1991
 APPLICATION NUMBER: US 08/039,320
 FILING DATE: 04-APR-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/403,576
 FILING DATE: 14-MAR-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Terskin, Robin L.
 REGISTRATION NUMBER: 35,030
 REFERENCE/DOCKET NUMBER: 017753-025
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 836-6620
 TELEFAX: (703) 836-2021
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2035 amino acids
 TYPE: amino acid
 STRANGENESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 FEATURE:
 NAME/KEY: Peptide
 LOCATION: 128_1899
 OTHER INFORMATION: /note= "The amino acids spanning 128 to 1899 constitute a repeated region wherein the repeat 1 OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such OTHER INFORMATION: repeats varies from 1 to 40."
 FEATURE:
 NAME/KEY: Peptide
 LOCATION: 134
 OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa
 Xaa Xaa which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."
 FEATURE:
 NAME/KEY: Peptide
 LOCATION: 144
 OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa
 which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACP
 or ACG; and Asn = AAT or AAC."
 FEATURE:
 NAME/KEY: Peptide
 LOCATION: 147
 OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa
 which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCP
 or CCG; and Ala = GCT, GCC, GCA, or GCG."
 FEATURE:
 NAME/KEY: Peptide
 LOCATION: 1..21

; OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
; 21 amino acid precursor sequence."
; US-09-134-916A-2

Query Match 100.0%; Score 49; DB 4; Length 2035;
 Best Local Similarity 100.0%; Pred. No. 4.4;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 STAPPVHN 9
 Db 1730 STAPPVHN 1738

RESULT 7
 US-08-787-547-55
 Sequence 55, Application US/08787547
 Patent No. 5783567
 GENERAL INFORMATION:
 APPLICANT: Hedley, Mary Lynne
 APPLICANT: Curley, Joanne M.
 APPLICANT: Lancer, Robert S.
 TITLE OF INVENTION: MICROPARTICLES FOR DELIVERY
 TITLE OF INVENTION: OF NUCLEIC ACID
 NUMBER OF SEQUENCES: 107
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Fish & Richardson, P.C.
 STREET: 225 Franklin Street
 CITY: Boston
 STATE: MA
 COUNTRY: US
 ZIP: 02110-2804
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: Windows95
 SOFTWARE: FastSEQ for Windows Version 2.0
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: US/08/787,547
 FILING DATE: 22-FAN-1997
 CLASSIFICATION: 514
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Fraser, Janis K.
 REGISTRATION NUMBER: 34,819
 REFERENCE/DOCKET NUMBER: 08191/003001
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 617-542-5070
 TELEFAX: 617-542-8906
 TELEX: 200154
 INFORMATION FOR SEQ ID NO: 55:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 9 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 US-08-787-547-55

Query Match 79.6%; Score 39; DB 1; Length 9;
 Best Local Similarity 77.8%; Pred. No. 3e+05;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 STAPPVHN 9
 Db 1 STAPPAGV 9

RESULT 8
 US-08-288-059-19
 Sequence 19, Application US/08288059
 Patent No. 5827666
 GENERAL INFORMATION:

APPLICANT: FINN, OLIVERA J.
 APPLICANT: FONTENOT, J. D.
 APPLICANT: MONTELARO, RONALD C.
 TITLE OF INVENTION: SYNTHETIC MULTIPLE TANDEM REPEAT MUCIN
 TITLE OF INVENTION: AND MUCIN-LIKE PEPTIDES, AND USES THEREOF
 NUMBER OF SEQUENCES: 36
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: CUSHMAN, DARBY & CUSHMAN, L.L.P.
 STREET: 1100 NEW YORK AVENUE, N.W.
 CITY: WASHINGTON
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20005

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/288,059
 FILING DATE: 08-AUG-1994
 CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: CHAPIN, MARLANA K.
 REGISTRATION NUMBER: 35,843

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-851-3711
 TELEFAX: 202-822-0944
 TELEX: 6714627 CUSH

INFORMATION FOR SEQ ID NO: 19:

SEQUENCE CHARACTERISTICS:

LENGTH: 9 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 US-08-288-059-19

Query Match Similarity 79.6%; Score 39; DB 2; Length 9;

Best Local Similarity 77.8%; Pred. No. 3e+05; Mismatches 0; Indels 0; Gaps 0;

Qy 1 STAPPYHNV 9
 Db 1 STAPPAGHV 9

RESULT 9
 US-09-593-870A-45
 i Sequence 45. Application US/09593870A
 i Patent No. 6548643

GENERAL INFORMATION:
 APPLICANT: McKenzie, Ian F.C.
 APPLICANT: Apostolopoulos, Vassos
 APPLICANT: Pietersz, Geoff Allian

TITLE OF INVENTION: Antigen Carbohydrate Compounds and Their
 FILE REFERENCE: 2368-MCKENZIE
 CURRENT APPLICATION NUMBER: US/09/593,870A

CURRENT FILING DATE: 2000-06-14
 PRIOR APPLICATION NUMBER: 09/223,043
 PRIOR FILING DATE: 1998-12-30

NUMBER OF SEQ ID NOS: 69
 SEQ ID NO: 45
 LENGTH: 9
 TYPE: PRT
 ORGANISM: Homo sapiens

US-09-593-870A-45
 Query Match Similarity 79.6%; Score 39; DB 4; Length 9;

Best Local Similarity 77.8%; Pred. No. 3e+05; Mismatches 2; Indels 0; Gaps 0;

Qy 1 STAPPYHNV 9
 Db 1 STAPPAGHV 9

RESULT 11
 US-09-043-731-19
 i Sequence 19. Application US/09043731A
 i Patent No. 6344203

GENERAL INFORMATION:
 APPLICANT: The Austin Research Institute
 TITLE OF INVENTION: Mimicking Peptides in Cancer Therapy

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Qy 1 STAPPYHNV 9
 Db 1 STAPPAGHV 9

RESULT 10
 US-09-497-232-1
 i Sequence 1. Application US/09497232
 i Patent No. 6600012

GENERAL INFORMATION:
 APPLICANT: AGRAWAL, Babita
 KRANTZ, Mark J.
 REDDISH, Mark A.
 LONGENECKER, B. Michael

TITLE OF INVENTION: METHOD FOR GENERATING ACTIVATED T-CELLS
 AND ANTIGEN-PULSED PRESENTING CELLS

NUMBER OF SEQUENCES: 34

CORRESPONDENCE ADDRESS:
 ADDRESSEE: FOLEY & LARDNER
 STREET: 3000 K Street, N.W.
 CITY: Washington
 STATE: D.C.
 COUNTRY: U.S.A.
 ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PC-DOS/MS-DOS
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/497,232
 FILING DATE: 03-FEB-2000
 CLASSIFICATION: <Unknown>
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: US/09/074,410
 FILING DATE: 08-MAY-1998
 APPLICATION NUMBER: US/09/497,232
 FILING DATE: 08-MAY-1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Saxe, Bernhard D.
 REGISTRATION NUMBER: 28,665
 REFERENCE/DOCKET NUMBER: 042881/0114
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 672-5300
 TELEFAX: (202) 672-5399
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 9 amino acids
 STRANDEDNESS: <Unknown>
 TOPOLOGY: Linear
 MOLECULE TYPE: Peptide
 SEQUENCE DESCRIPTION: SEQ ID NO: 1:
 US-09-497-232-1

Query Match Similarity 79.6%; Score 39; DB 4; Length 9;
 Best Local Similarity 77.8%; Pred. No. 3e+05; Mismatches 2; Indels 0; Gaps 0;

Qy 1 STAPPYHNV 9
 Db 1 STAPPAGHV 9

RESULT 9
 US-09-497-232-1
 i Sequence 9. Application US/09497232
 i Patent No. 6600012

GENERAL INFORMATION:
 APPLICANT: McKenzie, Ian F.C.
 APPLICANT: Apostolopoulos, Vassos
 APPLICANT: Pietersz, Geoff Allian

TITLE OF INVENTION: Use in Immunotherapy
 CURRENT APPLICATION NUMBER: US/09/593,870A

CURRENT FILING DATE: 2000-06-14
 PRIOR APPLICATION NUMBER: 09/223,043
 PRIOR FILING DATE: 1998-12-30

NUMBER OF SEQ ID NOS: 69
 SEQ ID NO: 45
 LENGTH: 9
 TYPE: PRT
 ORGANISM: Homo sapiens

US-09-593-870A-45
 Query Match Similarity 79.6%; Score 39; DB 4; Length 9;

FILE REFERENCE: CALA-200
 CURRENT APPLICATION NUMBER: US/09/043,731A
 NUMBER OF SEQ ID NOS: 26
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO: 19
 LENGTH: 16
 TYPE: PRT
 ORGANISM: Artificial Sequence
 FEATURE: Description of Artificial Sequence: single
 OTHER INFORMATION: stranded linear peptide
 US-09-043-731-19

Query Match 79.6%; Score 39; DB 4; Length 16;
 Best Local Similarity 77.8%; Pred. No. 1.3;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 CITY: WASHINGTON
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20005

RESULT 12
 US-09-049-354-3 Application US/08099354
 Patent No. 5344144

GENERAL INFORMATION:
 APPLICANT: FINN, OLIVERA J.
 FONTENOT, J. D.
 MONTELARO, RONALD C.
 TITLE OF INVENTION: SYNTHETIC MULTIPLE TANDEM REPEAT MUCIN
 AND MUCIN-LIKE PEPTIDES, AND USES THEREOF
 NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:
 ADDRESSEE: CUSHMAN, DARBY & CUSHMAN
 STREET: 1100 NEW YORK AVENUE, N.W.
 CITY: WASHINGTON
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20005

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/288,059-9
 FILING DATE: 08-AUG-1994
 CLASSIFICATION: 424
 ATTORNEY/AGENT INFORMATION:
 NAME: CHAPIN, MARLANA K.
 REGISTRATION NUMBER: 35,843
 REFERENCE/DOCKET NUMBER: 61137/205204

TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-822-0944
 TELEX: 671467 CUSH
 INFORMATION FOR SEQ ID NO: 9:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 19 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: Linear
 MOLECULE TYPE: peptide

US-08-288-059-9

Query Match 79.6%; Score 39; DB 2; Length 19;
 Best Local Similarity 77.8%; Pred. No. 1.6;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 STAPPVHN 9
 Db 8 STAPPVHGV 16

RESULT 14
 US-08-288-059-1 Application US/08288059
 Sequence 1, Application US/08288059
 Patent No. 5827666

GENERAL INFORMATION:
 APPLICANT: FINN, OLIVERA J.
 FONTENOT, J. D.
 MONTELARO, RONALD C.
 TITLE OF INVENTION: SYNTHETIC MULTIPLE TANDEM REPEAT MUCIN
 AND MUCIN-LIKE PEPTIDES, AND USES THEREOF
 NUMBER OF SEQUENCES: 36

CORRESPONDENCE ADDRESS:
 ADDRESSEE: CUSHMAN, DARBY & CUSHMAN, L.L.P.
 STREET: 1100 NEW YORK AVENUE, N.W.
 CITY: WASHINGTON
 STATE: D.C.
 COUNTRY: USA

US-08-099-354-3

Query Match 79.6%; Score 39; DB 1; Length 19;
 Best Local Similarity 77.8%; Pred. No. 1.6;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 STAPPVHN 9
 Db 8 STAPPVHGV 16

ZIP: 20005
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/288,059
 FILING DATE: 08-AUG-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: CHAPIN, MARLANA K.
 REGISTRATION NUMBER: 35,843
 REFERENCE/DOCKET NUMBER: 61137/205204
 TELEPHONE: 202-861-3711
 TELEFAX: 202-822-0944
 TELEX: 6714627 CUSH
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 20 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 US-08-288-059-1

Query Match 79.6%; Score 39; DB 2; Length 20;
 Best Local Similarity 77.8%; Pred. No. 1.7;
 Matches 0; Mismatches 2; Indels 0; Gaps 0;
 Qy 1 STAPPVHNV 9
 Db 9 STAPPAHGV 17

RESULT 15
 US-08-288-059-32
 Sequence 32, Application US/08288059
 Patent No. 5527666
 GENERAL INFORMATION:
 INN, OLIVERA J.
 APPLICANT: FONTENOT, J. D.
 APPLICANT: MONTELARO, RONALD C.
 TITLE OF INVENTION: SYNTHETIC MULTIPLE TANDEM REPEAT MUCIN
 TITLE OF INVENTION: AND MUCIN-LIKE PEPTIDES, AND USES THEREOF
 NUMBER OF SEQUENCES: 36
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
 STREET: 1100 NEW YORK AVENUE, N.W.
 CITY: WASHINGTON
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20005
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/288,059
 FILING DATE: 08-AUG-1994
 CLASSIFICATION: 424
 ATTORNEY/AGENT INFORMATION:
 NAME: CHAPIN, MARLANA K.
 REGISTRATION NUMBER: 35,843
 REFERENCE/DOCKET NUMBER: 61137/205204
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-861-3711
 TELEFAX: 202-822-0944
 INFORMATION FOR SEQ ID NO: 32:
 SEQUENCE CHARACTERISTICS:

XX Sequence 9 AA;
 SQ Sequence 9 AA;
 Query Match 100.0%; Score 49; DB 4; Length 9;
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 STAPPVHNV 9
 Db 1 STAPPVHNV 9

RESULT 2
 ABG79089 standard; peptide; 9 AA.
 ID ABG79089;
 AC ABG79089;
 XX DT 15-NOV-2002 (first entry)
 XX DB Human MUC1 class I HLA widely expressed antigen peptide #2.
 XX Cell penetrating peptide; cancer; tumour; melanoma; thymoma; antigen;
 KW lymphoma; sarcoma; lung cancer; non-Hodgkin's lymphoma; leukaemia;
 KW Hodgkin's lymphoma; uterine cancer; cervical cancer; bladder cancer;
 KW kidney cancer; adenocarcinoma; breast cancer; prostate cancer;
 KW ovarian cancer; pancreatic cancer; epoxide; vaccine; dendritic cell;
 KW tumour infiltrating lymphocyte; TIL; human leukocyte antigen; HLA;
 KW cytostatic; human.
 OS Homo sapiens.
 XX PN WO2000064057-A2.
 XX PD 22-AUG-2002.
 XX PF 15-FEB-2002; 2002WO-US005212.
 XX PR 15-FEB-2001; 2001US-0268687P.
 XX PA (BAYU) BAYLOR COLLEGE MEDICINE.
 XX PI Wang R;
 XX DR WPI; 2002-627577/67.
 XX Novel composition for treating a disease in an animal, comprises an
 PT immune effector cell and cell penetrating peptide associated with an
 PT antigen or antibody.
 XX Disclosure; Page 18; 61pp; English.

XX The invention relates to a composition (I) comprising an immune effector cell and a cell penetrating peptide (CPP) associated with an antigen or antibody. Also included are (1) a vaccine comprising (I), CPP associated with an antigen, and a pharmaceutically acceptable carrier and (2) preparing a composition for a disease, by providing (I) and CPP associated with an antigen for disease, and introducing the antigen-associated CPP to (I), where antigen enters into the cell. The antigens are, for example, tumour antigen derived epitopes recognised by tumour infiltrating lymphocytes (TIL) of HLA (human leukocyte antigen) class I or II. The composition is useful for enhancing immunity in an animal to a disease, by administering a mature dendritic cell comprising CPP associated with an antigen to the animal, such that following administration, the animal is protected from disease, where the animal comprises both CD4+ and CD8+ T cells. It is also useful for treating a disease (e.g., cancer, tumour, melanoma, thymoma, lymphoma, sarcoma, lung cancer, non-Hodgkin's lymphoma, leukaemia, Hodgkin's lymphoma, uterine cancer, cervical cancer, bladder cancer, kidney cancer, adenocarcinoma, breast cancer, prostate cancer, ovarian cancer and pancreatic cancer).
 CC The animal is further subjected to a cancer treatment including surgery, radiation, chemotherapy or gene therapy. The administration of (I), preferably dendritic cell is prior to, subsequent to or concurrent with,

CC the cancer treatment. The present sequence is a tumour antigen derived epitope for inclusion in the composition of the invention.

XX Sequence 9 AA;
 SQ Sequence 9 AA;
 Query Match 100.0%; Score 49; DB 5; Length 9;
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 STAPPVHNV 9
 Db 1 STAPPVHNV 9

RESULT 3
 ADA50588
 ID ADA50588 standard; peptide; 9 AA.
 XX AC ADA50588;
 XX DT 20-NOV-2003 (first entry)
 XX DE Mucin 1 (MUC-1) CTL epitope, SEQ ID NO:43.
 XX Nucleic acid vaccine; DNA vaccine; tumour antigen; cytokine adjuvant; humoral response; cellular response; immune response; immunotherapy; cancer; cytotoxic; vaccine; gene therapy; mucin 1; MUC-1; cytotoxic T lymphocyte; CTL epitope.
 KW Unidentified.
 OS XX
 XX PN WO2003031569-A2.
 XX PD 17-APR-2003.
 XX PF 18-SEP-2002; 2002WO-US029640.
 XX PR 10-OCT-2001; 2001US-0328371P.
 XX PA (CENZ) CENTOCOR INC.
 XX PI Snyder L, Scallan B, Knight DM, McCarthy SG, Goletz TJ;
 PI Branigan PJ;
 XX DR WPI; 2003-393437/37.
 XX New nucleic acid vaccine, useful for eliciting an immune response to a cancer associated tumor protein in a mammal.
 XX Claim 1a; Page 45; 92pp; English.

XX The invention relates to a nucleic acid vaccine comprising one or more tumour antigen-encoding nucleic acids and one or more cytokine adjuvant-encoding nucleic acids. The tumour antigen encoded by the vaccine is mucin 1 (MUC-1), the kallikrein Kk2, or prostate specific antigen (PSA, also known as Klk3), and the cytokine adjuvant encoded can be interleukin -12 (IL-12), granulocyte macrophage-colony stimulating factor (GM-CSF), or especially interleukin-18 (IL-18). The antigen-encoding nucleic acid is preferably under the control of a promoter such as the cytomegalovirus immediate early promoter, the dihydrofolate reductase promoter or the early or late SV40 promoters. The invention also encompasses the method of eliciting an immune response to a tumour antigen in a mammal using the vaccine of the invention. Coexpression of the antigen and adjuvant induces a humoral or cellular response to the tumour antigen, generating an immune response useful for treatment or prophylaxis of cancers. The present sequence represents a mucin 1 (MUC-1) polypeptide sequence which is specifically claimed for use in the vaccine of the invention.

XX Sequence 9 AA;
 SQ Sequence 9 AA;
 Query Match 100.0%; Score 49; DB 6; Length 9;
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1 STAPPVHNV 9	KW	viral infection; human immunodeficiency virus; melanoma;
Db	1 STAPPVHNV 9	KW	bacterial infection; Salmonella; Legionella; Giardia.
		XX	Trypanosoma; Toxoplasma; Giardia.
		OS	Homo sapiens.
		XX	XX
		PN	W0200190197-A1.
		XX	XX
		PD	29-NOV-2001.
		XX	XX
		AC	25-MAY-2001; 2001WO-AU000622.
		XX	XX
		DT	26-MAY-2000; 2000AU-00007761.
		XX	XX
		DE	PA (AUSU) UNIV AUSTRALIAN NAT.
		XX	XX
		KW	New synthetic polypeptides having several different segments of at least
		XX	PI Thomson SA, Ramshaw IA;
		XX	XX
		KW	PT New synthetic polypeptides having several different segments of at least
		XX	PR one parent polypeptide linked together differently compared to the
		XX	PR linkage in the parent polypeptide, for inducing immune response against a
		XX	PR pathogen or cancer.
		OS	XX
		XX	PS Example 3; Fig 27; 364pp; English.
		PN	XX
		XX	CC The invention relates to a new synthetic polypeptide (I) comprising
		W09837095-A2.	CC several different segments of at least one parent polypeptide linked
		XX	CC together in a different relationship relative to their linkage in the
		PD	CC parent polypeptide to impede, abrogate or otherwise alter at least one
		27-AUG-1998.	CC function associated with the parent polypeptide and for inducing an
		XX	CC immune response against a pathogen or cancer. Also included are a
		PF	CC synthetic polynucleotide encoding and a computer system for designing the
		24-FEB-1998;	CC synthetic polypeptides. The synthetic polypeptide is useful for
		XX	CC modulating immune responses preferably directed against a pathogen or a
		PR	CC cancer, (e.g., cancers of the lung, breast, ovary, cervix, colon, head,
		24-FEB-1997;	CC and neck, pancreas, prostate, bladder, kidney, bone liver,
		XX	CC oesophagus, brain, testicle, uterus), as potentiating agents.
		PA	CC Compositions comprising the polypeptide may be used in the treatment or
		(THER-) THERION BIOLOGICS CORP.	CC prophylaxis against viral (such as infections caused by HIV (human
		PA	CC immunodeficiency virus), hepatitis, influenza, Japanese encephalitis
		(USSH) US DEPT HEALTH & HUMAN SERVICES.	CC virus, Epstein-Barr virus and respiratory syncytial virus), bacterial
		PA	CC (e.g., infections caused by Neisseria, Meningococcal, Haemophilus
		(DAND) DANA FARBER CANCER INST INC.	CC Salmonella, Streptococcal, Legionella, and Mycobacterium or parasitic
		XX	CC (e.g., infections caused by Plasmodium, Schistosoma, Leishmania,
		PI	CC Trypanosoma, Toxoplasma and Giardia) infections. The present sequence is
		Schlom J, Kantor J, Kufe D, Panicali D, Gritz L;	CC a peptide derived from a parent protein used to construct a savine of the
		XX	CC invention
		DR	XX
		1998-467492/40.	SQ Sequence 30 AA;
		XX	Query Match 100.0%; Score 49; DB 2; Length 13;
		PT	Best Local Similarity 100.0%; Pred. No. 0.11;
		PT	Mismatches 0; Indels 0; Gaps 0;
		Example 1; Page 20; 42pp; English.	Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
		XX	Qy 1 STAPPVHNV 9
		PS	Db 2 STAPPVHNV 10
		XX	XX
		XX	RESULT 5
		AAU84987	AAU84987
		ID AAU84987 standard; peptide; 30 AA.	ID AAY71021 standard; protein; 173 AA.
		XX	XX
		AC AAU84987;	AC AAY71021;
		XX	XX
		DT 08-MAY-2002 (first entry)	DT 29-AUG-2000 (first entry)
		XX	XX
		DE Human MUC1R segment 1.	DE Human Mucin 1 (MUC-1) protein fragment #2.
		XX	XX
		KW Savine; vaccine; cancer; viral infection; HIV; hepatitis C virus;	

KW Human; Mucin 1; MUC-1; tumour; pMRS30 expression vector; anti-tumour;
 KW therapy; immune response; cytostatic; vaccine.
 XX
 OS Homo sapiens.
 XX
 PN WO200025827-A2.
 XX
 PD 11-MAY-2000.
 XX
 PP 18-OCT-1999; 99WO-BP007874.
 XX
 PR 30-OCT-1998; 98IT-MI002330.
 XX
 PA (MENA) MENARINI RICERCHE SPA.
 XX
 PI Parente D, Di Massimo AM, De Santis R;
 XX
 DR WPI; 2000-365410/31.
 DR N-PSDB; AAD00385.
 XX
 PT Composition containing one or more DNA molecules encoding fragments of a
 PT Mucin 1 (MUC-1) protein overexpressed in tumor cells, useful in anti-
 PT tumor therapy.
 XX
 PS Claim 16; Fig 2; 56pp; English.
 XX
 CC The present sequence is a fragment of human Mucin 1 (MUC-1), an antigenic
 CC protein overexpressed in tumour cells. The sequence was obtained from
 CC BT20 tumour cells. The corresponding DNA sequence is cloned into a pMRS30
 CC expression vector and used in pharmaceutical composition e.g. vaccine for
 CC inducing an antigen-specific anti-tumour immune response. Composition
 CC containing this DNA molecule is useful in anti-tumour therapy of patients
 CC affected with tumours characterised by high MUC-1 expression
 XX Sequence 173 AA;
 SQ Score 100.0%; Best Local Similarity 100.0%; Pred. No. 1.5;
 SQ Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 STAPPVHN 9
 Db 127 STAPPVHN 135
 XX
 RESULT 7
 ID AAR27664
 XX Standard; protein; 180 AA.
 AC AAR27664;
 XX
 DT 25-MAR-2003 (revised)
 DT 06-NOV-1992 (first entry)
 XX
 DE C-terminal region of H23-ETA-S antigen.
 KW Secreted; human epithelial antigen; Monoclonal antibody H23; vaccine;
 KW malignant tumour; breast cancer; tandem repeat.
 XX
 OS Homo sapiens.
 XX
 FH Key Region 1..40 Location/Qualifiers
 FT /note= "contains 2 tandem repeats - can have up to 80 copies"
 FT Misc-difference 7
 FT /label= Pro, Ala
 FT /note= "natural polymorphism"
 FT Misc-difference 17
 FT /label= Thr, Asn
 FT /note= "natural polymorphism"
 FT Misc-difference 20
 FT /label= Pro, Ala
 FT
 FT Misc-difference 27
 FT /label= PRO, Ala
 FT Misc-difference 37
 FT /label= Thr, Asn
 FT Misc-difference 40
 FT /label= PRO, Ala
 FT /note= "natural polymorphism"
 XX
 PN WO9207000-A1.
 XX
 PD 30-APR-1992.
 XX
 PI 91WO-FR000835.
 XX
 PR 23-OCT-1990; 90FR-00013101.
 XX
 PA (TRGE) TRANSGENE SA.
 XX
 PT Compds. contg. polypeptide antigen recognised by antibody H23 - for
 PT treatment of mammary tumours, also for pox virus compns. for use in
 PT vaccines.
 XX
 DR WPI; 1992-167097/20.
 DR N-PSDB; AAQ24681.
 XX
 PA The tumour antigen recognised by antibody H23 is aberrantly expressed in
 CC epithelial cells from cancerous mammary tissue in about 90 per cent of
 CC breast cancer cases; in a normal individual expression is negligible. The
 CC antigen exists in two forms: transmembrane (ETA-T) and secreted (ETA-S).
 CC Both forms show high degree of polymorphism. A 20 amino acid subunit in
 CC ETA can be tandemly repeated up to 80 times. (N.B. two tandem repeats are
 CC shown here; the first half of the amino acid sequence, i.e. on the N-
 CC terminal side of the repeat region, is given in AAR27663). From one
 CC subunit to the next, 1 to 3 amino acids can differ. See also AAQ24678-
 CC Q24681, AAQ29276-7 and AAR23974-R23981. (Updated on 25-MAR-2003 to
 CC correct PN field.)
 XX
 SQ Sequence 180 AA;
 Query Match 100.0%; Best Local Similarity 100.0%; Pred. No. 1.6;
 SQ Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 STAPPVHN 9
 Db 43 STAPPVHN 51
 XX
 RESULT 8
 ID AAR27665
 XX Standard; protein; 287 AA.
 AC AAR27665;
 XX
 DT 25-MAR-2003 (revised)
 DT 06-NOV-1992 (first entry)
 XX
 DE Secreted form of H23-ETA antigen.
 KW Homo sapiens.
 XX
 FH Key Region 1..40 Location/Qualifiers
 FT /note= "contains 2 tandem repeats - can have up to 80 copies"
 FT Misc-difference 7
 FT /label= Pro, Ala
 FT /note= "natural polymorphism"
 FT Misc-difference 17
 FT /label= Thr, Asn
 FT /note= "natural polymorphism"
 FT Misc-difference 20
 FT /label= Pro, Ala
 FT
 XX
 PH Key Peptide
 XX
 PH Location/Qualifiers
 1..21

Copyright GenCore version 5.1.6 (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 28, 2004, 18:10:44 ; Search time 42 Seconds (without alignments)

67.218 Million cell updates/sec

Title: US-10-019-513-1

Perfect score: 49

Sequence: 1 STAPPVHN 9

Scoozing table: BLOSUM62 Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0% Maximum Match 100% Listing First 45 summaries

Database : Published Applications AA:^{*}

1: /cgna2_6/ptodata/1/pubpaas/US07_PUBCOMB.pep:*

2: /cgna2_6/ptodata/1/pubpaas/US08_PUBCOMB.pep:*

3: /cgna2_6/ptodata/1/pubpaas/US09_PUBCOMB.pep:*

4: /cgna2_6/ptodata/1/pubpaas/US10_PUBCOMB.pep:*

5: /cgna2_6/ptodata/1/pubpaas/US11_PUBCOMB.pep:*

6: /cgna2_6/ptodata/1/pubpaas/US12_PUBCOMB.pep:*

7: /cgna2_6/ptodata/1/pubpaas/US13_PUBCOMB.pep:*

8: /cgna2_6/ptodata/1/pubpaas/US14_PUBCOMB.pep:*

9: /cgna2_6/ptodata/1/pubpaas/US15_PUBCOMB.pep:*

10: /cgna2_6/ptodata/1/pubpaas/US16_PUBCOMB.pep:*

11: /cgna2_6/ptodata/1/pubpaas/US09C_PUBCOMB.pep:*

12: /cgna2_6/ptodata/1/pubpaas/US09_NEW_PUBCOMB.pep:*

13: /cgna2_6/ptodata/1/pubpaas/US10A_PUBCOMB.pep:*

14: /cgna2_6/ptodata/1/pubpaas/US10B_PUBCOMB.pep:*

15: /cgna2_6/ptodata/1/pubpaas/US10C_PUBCOMB.pep:*

16: /cgna2_6/ptodata/1/pubpaas/US10_NEW_PUBCOMB.pep:*

17: /cgna2_6/ptodata/1/pubpaas/US60_NEW_PUBCOMB.pep:*

18: /cgna2_6/ptodata/1/pubpaas/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	Match Length	DB ID	Description
1	49	100.0	9	12	US-10-247-703-43
2	49	100.0	9	16	US-10-447-162-98
3	49	100.0	13	14	US-10-057-136-18
4	49	100.0	30	12	US-10-296-734-1168
5	49	100.0	307	12	US-10-247-703-26
6	49	100.0	312	12	US-10-296-734-824
7	49	100.0	475	15	US-10-417-312-21
8	49	100.0	508	14	US-10-057-136-20
9	49	100.0	515	12	US-10-612-030-19
10	49	100.0	515	12	US-10-247-703-20
11	49	100.0	515	12	US-10-247-703-19
12	49	100.0	515	14	US-10-097-340-212
13	49	100.0	515	14	US-10-171-311-156
14	49	100.0	1255	10	US-09-996-699-10
15	49	100.0	1255	14	US-10-171-311-158

* Query Match Score 49; DB ID US-10-247-703-43

Sequence 43, App1 Sequence 98, App1 Sequence 18, App1 Sequence 1168, App1 Sequence 26, App1 Sequence 824, App1 Sequence 22, App1 Sequence 1, App1 Sequence 20, App1 Sequence 19, App1 Sequence 12, App1 Sequence 21, App1 Sequence 156, App1 Sequence 10, App1 Sequence 158, App1

Sequence 311, App1 Sequence 1210, App1 Sequence 861, App1 Sequence 55, App1 Sequence 44, App1 Sequence 45, App1 Sequence 57, App1 Sequence 97, App1 Sequence 47, App1 Sequence 138, App1 Sequence 11, App1 Sequence 1, App1

Sequence 3, App1 Sequence 1, App1 Sequence 1, App1 Sequence 16, App1 Sequence 17, App1 Sequence 20, App1 Sequence 40, App1 Sequence 31, App1 Sequence 215, App1 Sequence 32, App1 Sequence 196, App1 Sequence 207, App1 Sequence 1, App1 Sequence 63, App1 Sequence 5, App1

ALIGMENTS

RESULT 1 ; Sequence 43, Application US/10247703-43 ; Publication No. US2003006597A1 ; GENERAL INFORMATION: ; APPLICANT: Branigan, Patrick ; INVENTOR: Goletz, Theresa J ; APPLICANT: Knight, David M ; APPLICANT: McCarthy, Stephen G ; APPLICANT: Scallion, Bernard J ; APPLICANT: Snyder, Linda A ; TITLE OF INVENTION: NUCLEIC ACID VACCINES USING TUMOR ANTIGEN ENCODING NUCLEIC ACIDS ; FILE REFERENCE: CEN310 ; CURRENT APPLICATION NUMBER: US/10/247-703-43 ; CURRENT FILING DATE: 2002-09-20 ; PRIORITY APPLICATION NUMBER: 60/328,371 ; PRIORITY FILING DATE: 2001-10-10 ; NUMBER OF SEQ ID NOS: 77 ; SOFTWARE: PatentIn version 3.1 ; SEQ ID NO 43

Query Match Score 100.0%; Pred. No. 1.2e+06; DB 12; Length 9; ; TYPE: PRT ; ORGANISM: Homo sapiens ; Best Local Similarity 100.0%; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 STAPPINV 9 ||| ||||| ||| 1 STAPPINV 9

Db

RESULT 2 ; Sequence 98, Application US/10447161-98 ; Sequence 10, Application US/10447161

Publication No. US20040023314A1
 GENERAL INFORMATION:
 APPLICANT: Wang, Rong-fu
 TITLE OF INVENTION: Mutant Fibronectin and Tumor Metastasis
 FILE REFERENCE: HO-P02484US1
 CURRENT APPLICATION NUMBER: US/10/447,161
 CURRENT FILING DATE: 2003-05-28
 PRIOR APPLICATION NUMBER: 60/383,530
 PRIOR FILING DATE: 2002-05-28
 NUMBER OF SEQ ID NOS: 148
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO: 98
 LENGTH: 9
 TYPE: PRT
 ORGANISM: Artificial Sequence
 OTHER INFORMATION: Synthetic Peptide
 FEATURE: OTHER INFORMATION: Synthetic Peptide
 SEQ ID NO: 10-447-161-98
 Query Match 100.0%; Score 49; DB 16; Length 9;
 Best Local Similarity 100.0%; Pred. No. 1.2e+06;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 STAPPVHN 9
 Db 1 STAPPVHN 9
 RESULT 3
 US-10-57-136-18
 Sequence 18, Application US/10057136
 Publication No. US20030021770A1
 GENERAL INFORMATION:
 APPLICANT: SCHLOM, JEFFREY
 APPLICANT: KANTOR, JUDITH
 APPLICANT: KUPE, DONALD
 APPLICANT: PANICALI, DENNIS
 APPLICANT: GRITZ, LINDA
 TITLE OF INVENTION: RECOMBINANT POX VIRUS FOR IMMUNIZATION AGAINST MUCC1
 FILE REFERENCE: 700953/47113C
 CURRENT APPLICATION NUMBER: US/10/057,136
 CURRENT FILING DATE: 2002-01-25
 PRIOR APPLICATION NUMBER: 09/366,670
 PRIOR FILING DATE: 1999-08-03
 PRIOR APPLICATION NUMBER: PCT/US98/03693
 PRIOR FILING DATE: 1998-02-24
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO: 18
 LENGTH: 13
 TYPE: PRT
 ORGANISM: Homo sapiens
 SEQ ID NO: 10-057-136-18
 Query Match 100.0%; Score 49; DB 14; Length 13;
 Best Local Similarity 100.0%; Pred. No. 0.1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 STAPPVHN 9
 Db 2 STAPPVHN 10
 RESULT 4
 US-10-296-734-1168
 Sequence 1168, Application US/10296734
 Publication No. US20040054137A1
 GENERAL INFORMATION:
 APPLICANT: Thompson, Scott A
 APPLICANT: Thompson, Ian A
 TITLE OF INVENTION: Synthetic molecules and uses therefor
 FILE REFERENCE: Savine
 Publication No. US20040054137A1
 CURRENT APPLICATION NUMBER: US/10/296,734
 CURRENT FILING DATE: 2003-08-04
 PRIOR APPLICATION NUMBER: AU PQ7761/00
 PRIOR FILING DATE: 2000-05-26

NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO: 824
; LENGTH: 312
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: MUC1R consensus polypeptide
US-10-296-734-824

Query Match 100.0%; Score 49; DB 12; Length 312;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 STAPPVHN 9
Db 7 STAPPVHN 15

RESULT 7
US-10-247-703-22
; Sequence 22, Application US/10247703
; Publication No. US2003006359A1
; GENERAL INFORMATION:
; APPLICANT: Branigan, Patrick
; APPLICANT: Goletz, Theresa J
; APPLICANT: Knight, David M
; APPLICANT: McCarthy, Stephen G
; APPLICANT: Scallion, Bernard J
; APPLICANT: Snyder, Linda A
; TITLE OF INVENTION: NUCLEIC ACID VACCINES USING TUMOR ANTIGEN ENCODING NUCLEIC ACIDS
; FILE REFERENCE: CEN10
; CURRENT APPLICATION NUMBER: US/10/247,703
; CURRENT FILING DATE: 2002-09-20
; PRIOR APPLICATION NUMBER: 60/328,371
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 22
; LENGTH: 475
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-247-703-22

Query Match 100.0%; Score 49; DB 12; Length 475;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 STAPPVHN 9
Db 170 STAPPVHN 178

Query Match 100.0%; Score 49; DB 15; Length 475;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 STAPPVHN 9
Db 170 STAPPVHN 178

RESULT 9
US-10-057-136-20
; Sequence 20, Application US/10057136
; Publication No. US20030021770A1
; GENERAL INFORMATION:
; APPLICANT: SCHIOM, JEFFREY
; APPLICANT: KANTOR, JUDITH
; APPLICANT: KUFE, DONALD
; APPLICANT: PANICALLI, DENNIS S
; APPLICANT: GRITZ, LINDA
; TITLE OF INVENTION: RECOMBINANT POX VIRUS FOR IMMUNIZATION AGAINST MUC1
; FILE REFERENCE: 700953/47113C
; CURRENT APPLICATION NUMBER: US/10/057,136
; CURRENT FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: 1999-08-03
; PRIOR FILING DATE: 1998-02-24
; PRIOR APPLICATION NUMBER: PCT/US98/03693
; PRIOR FILING DATE: 1998-02-24
; PRIOR APPLICATION NUMBER: 60/038,253
; PRIOR FILING DATE: 1997-02-24
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO: 20
; LENGTH: 508
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-057-136-20

Query Match 100.0%; Score 49; DB 14; Length 508;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 STAPPVHN 9
Db 203 STAPPVHN 211

RESULT 10
US-10-612-090-19
; Sequence 19, Application US/10612090
; Publication No. US20040057952A1
; GENERAL INFORMATION:
; APPLICANT: Immunogen, Inc.
; FILE REFERENCE: A8340
; CURRENT APPLICATION NUMBER: US/10/612,090
; CURRENT FILING DATE: 2003-07-03
; PRIOR APPLICATION NUMBER: US 60/393,094
; PRIOR FILING DATE: 2002-07-03
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO: 19
; LENGTH: 515
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Exemplary Muc1 protein
US-10-612-090-19

Query Match 100.0%; Score 49; DB 12; Length 515;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 STAPPVHN 9
Db 170 STAPPVHN 178

```

i PRIORITY APPLICATION NUMBER: 60/3/24, 367
i PRIORITY FILING DATE: 2001-09-26
i PRIORITY APPLICATION NUMBER: 60/3/11, 732
i PRIORITY FILING DATE: 2001-08-10
i PRIORITY APPLICATION NUMBER: 60/3/25, 102
i PRIORITY FILING DATE: 2001-09-26
i PRIORITY APPLICATION NUMBER: 60/3/23, 580
i PRIORITY FILING DATE: 2001-09-19
i PRIORITY APPLICATION NUMBER: 60/3/23, 580
; NUMBER OF SEQ ID NOS: 363
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO: 212
; LENGTH: 515
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-097-340-212

Query Match          100.0%; Score 49; DB 14; Length 515;
Best Local Similarity 100.0%; Pred. No. 3.4; Mismatches 0; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 11
; Sequence 20, Application US/10247703
; GENERAL INFORMATION:
; APPLICANT: Braniisan, Patrick
; APPLICANT: Golter, Theresa J
; APPLICANT: Knight, David M
; APPLICANT: McCarthy, Stephen G
; APPLICANT: McCallion, Bernard J
; APPLICANT: Snyder, Linda A
; TITLE OF INVENTION: NUCLEIC ACID VACCINES USING TUMOR ANTIGEN ENCODING NUCLERIC ACIDS
; FILE REFERENCE: CEN310
; CURRENT FILING DATE: 2002-09-10/247,703
; PRIORITY APPLICATION NUMBER: 60/328,371
; SEQ ID NO: 20
; LENGTH: 515
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-247-703-20

Query Match          100.0%; Score 49; DB 12; Length 515;
Best Local Similarity 100.0%; Pred. No. 3.4; Mismatches 0; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 12
; Sequence 212, Application US/10097340
; GENERAL INFORMATION:
; APPLICANT: John MONAHAN
; APPLICANT: Manjula GANNAVARAM
; APPLICANT: Sebastian HOERSCH
; APPLICANT: Shubhangi KAMATKAR
; APPLICANT: Steve G. KOVATS
; APPLICANT: Rachel E. MEYERS
; APPLICANT: Michael MORISEY
; APPLICANT: Peter OLANET
; APPLICANT: Ami SEN
; APPLICANT: Peter VEIBY
; APPLICANT: Gordon B. MILLS
; APPLICANT: Robert C. BAST, Jr.
; APPLICANT: Karen LU
; APPLICANT: Rosemarie SCHMANDT
; APPLICANT: Xumei ZHAO
; APPLICANT: Karen GLATT
; TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,
; TITLE OF INVENTION: Assessment, Prevention, and Therapy of Ovarian Cancer
; CURRENT APPLICATION NUMBER: US/10/097,340
; CURRENT FILING DATE: 2002-03-14
; PRIORITY APPLICATION NUMBER: 60/2/76, 025
; PRIOR FILING DATE: 2001-03-14
; PRIORITY APPLICATION NUMBER: 60/3/25, 149
; PRIOR FILING DATE: 2001-09-26
; PRIORITY APPLICATION NUMBER: 60/2/76, 026
; PRIOR FILING DATE: 2001-03-14

Query Match          100.0%; Score 49; DB 14; Length 515;
Best Local Similarity 100.0%; Pred. No. 3.4; Mismatches 0; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 13
; Sequence 156, Application US/10171311
; Publication No. US2003008720A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Giatt, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoersch, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; TITLE OF INVENTION: OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIORITY APPLICATION NUMBER: US 60/2/98,159
; PRIOR FILING DATE: 2001-06-13
; PRIORITY APPLICATION NUMBER: US 60/2/98,155
; PRIOR FILING DATE: 2001-06-13
; PRIORITY APPLICATION NUMBER: US 60/3/35,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO: 156
; LENGTH: 515
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-171-311-156

Query Match          100.0%; Score 49; DB 14; Length 515;
Best Local Similarity 100.0%; Pred. No. 3.4; Mismatches 0; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 14
; Sequence 10, Application US/09996069
; Publication No. US2003036199A1
; GENERAL INFORMATION:
; APPLICANT: Bamdad, Cynthia
; APPLICANT: Baabdad, Cynthia

```

```

; APPLICANT: Bamdad, R. Shoshana
; TITLE OF INVENTION: DIAGNOSTIC TUMOR MARKERS, DRUG SCREENING FOR TUMORIGENESIS INHIBI
; FILE REFERENCE: MOL015/70071
; CURRENT APPLICATION NUMBER: US/09/996, 069
; CURRENT FILING DATE: 2001-11-27
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 1255
; TYPE: PRT
; - ORGANISM: Homo sapiens
US-09-996-069-10

```

```

Query Match          100.0%; Score 49; DB 10; Length 1255;
Best Local Similarity 100.0%; Pred. No. 8.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy      1 STAPPVNV 9
Db      950 STAPPVNV 958

```

```

RESULT 15
US-10-171-311-158
Sequence 158, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Giatt, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoersh, Sebastian
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171-311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 158
; LENGTH: 1255
; TYPE: PRT
; - ORGANISM: Homo sapiens
US-10-171-311-158

```

```

Query Match          100.0%; Score 49; DB 14; Length 1255;
Best Local Similarity 100.0%; Pred. No. 8.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy      1 STAPPVNV 9
Db      950 STAPPVNV 958

```



FT Protein /label= signal
FT 22 . 287
FT /label= ETA-T
FT Misc-difference 134
FT /label= Pro, Ala
FT /note= "natural polymorphism"
FT Misc-difference 144
FT /label= Thr, Asn
FT /note= "natural polymorphism"
FT Misc-difference 147
FT /label= Pro, Ala
FT /note= "natural polymorphism"
XX WO9207000-A1.
XX 30-APR-1992.
XX 23-OCT-1991; 91WO-FR0000835.
XX 23-OCT-1990; 90FR-00013101.
PA (TIGE) TRANSGENE SA.
PI Chambon P, Kiery MP, Lathe R, Hareveni M;
DR WPI; 1992-167097/20.
XX Compns. consq. polypeptide antigen recognised by antibody H23 - for
PT treatment of mammary tumours, also for pox virus compsns. for use in
PT vaccines.
XX PS Claim 3; Page 19-21; 29pp; French.
XX The tumour antigen recognised by antibody H23 is aberrantly expressed in
CC epithelial cells from cancerous mammary tissue in about 90 per cent of
CC breast cancer cases; in a normal individual expression is negligible. The
CC antigen exists in two forms: transmembrane (ETA-T) and secreted (ETA-S).
CC Both forms show a high degree of polymorphism. A 20 amino acid subunit in
CC ETA can be tandemly repeated up to 80 times. From one subunit to the
CC next, 1 to 3 amino acids can differ. DNA coding for immunogenic fragments
CC of ETA can be inserted into e.g. vaccinia viruses for treatment of
CC mammary tumours. See also AAQ24678-Q24681, AAQ29276-7 and AAR23974-
XX R23981. (Updated on 25-MAR-2003 to correct PN field.)
SQ Sequence 287 AA;
Query Match Best Local Similarity 100.0%; Score 49; DB 2; Length 287;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 STAPPVHN 9
Db 150 STAPPVHN 158
RESULT 9
AYY71027 ID AYY71027 standard; protein; 295 AA.
XX AC AAY71027;
XX DT 12-SEP-2003 (revised)
DT 29-AUG-2000 (first entry)
DE Ubiquitin-E. coli LacI-human Mucin 1 fusion protein #2.
KW Ubiquitin; LacI; beta-galactosidase; fusion protein; human; Mucin 1;
KW MUC-1; tumour; pMRS30 expression vector; anti-tumour; therapy;
KW immune response; cytosstatic; vaccine.
OS Homo sapiens.
OS Escherichia coli.

OS XX Chimeric.
FH Key
FT Region 1 . 123
FT /label= UBILaci_protein
FT /note= "contains ubiquitin sequence fused to a portion of
E. coli LacI"
FT Region 124 . 295
FT /label= Human_MUC-1_fragment
PN WO200025827-A2.
XX PD 11-MAY-2000.
XX PP 18-OCT-1999;
XX DR WPI; 2000-365410/31.
XX N-PSDB; AAD00391.
XX PT Composition containing one or more DNA molecules encoding fragments of a
PT Mucin 1 (MUC-1) protein overexpressed in tumor cells, useful in anti-
PT tumor therapy.
XX PI Patente D, Di Massimo AM, De Santis R;
XX DR
XX PA (MENARINI RICERCHE SPA.
XX PT
XX PS Claim 1B; Fig 8; 56pp; English.
XX CC The present sequence is a fusion protein consisting of human Mucin 1 (MUC-1)
CC -1) fragment fused to UBILaci sequence at the N-terminus. The UBILaci
CC sequence consists of ubiquitin sequence from MCF7 cell line and a portion of E.
coli beta-galactosidase (lacI). MUC-1 is an antigenic protein
overexpressed in tumour cells. The corresponding DNA sequence is cloned
into a pMRS30 expression vector and used in pharmaceutical composition
e.g. vaccine for inducing an antigen-specific anti-tumour immune
response. Composition containing this DNA molecule is useful in anti-
CC tumour therapy of patients affected with tumours characterised by high
CC MUC-1 expression. (Updated on 12-SEP-2003 to standardise OS field)
XX Sequence 295 AA;
SQ Query Match 100.0%; Score 49; DB 3; Length 295;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 STAPPVHN 9
Db 249 STAPPVHN 257
RESULT 10
ADA50571 ID ADA50571 standard; protein; 307 AA.
XX AC ADA50571;
XX DT 20-NOV-2003 (first entry)
XX DE Mucin 1 (MUC-1) splice variant #1, SEQ ID NO:26.
XX KW Nucleic acid vaccine; DNA vaccine; tumour antigen; cytokine adjuvant;
KW humoral response; cellular response; immune response; immunotherapy;
KW cancer; cytosstatic; vaccine; gene therapy; mucin 1; MUC-1.
OS Unidentified.
XX FN WO2003031569-A2.
XX PD 17-APR-2003.
XX

PF 18-SEP-2002; 2002WO-US029640.
 XX DR WPI; 2002-147575/19.
 PR 10-OCT-2001; 2001US-0328371P.
 XX PT New synthetic polypeptides having several different segments of at least
 PA (CENZ) CENTOCOR INC.
 PT one parent polypeptide linked together differently compared to the
 PT linkage in the parent polypeptide, for inducing immune response against a
 PT pathogen or cancer.
 XX PI Snyder L, Scallion B, Knight DM, McCarthy SG, Goletz TJ,
 PI Branigan PJ;
 XX PS Example 3; Fig 27; 364pp; English.
 XX DR The invention relates to a new synthetic polypeptide (I) comprising
 PT several different segments of at least one parent polypeptide linked
 PT together in different relationship relative to their linkage in the
 PT parent polypeptide to impede, abrogate or otherwise alter at least one
 PT linkage associated with the parent polypeptide and for inducing an
 XX function associated with a pathogen or cancer. Also included are:
 CC a computer system for designing the
 CC synthetic polypeptides and poly nucleotides
 CC encoding the synthetic polypeptides. The synthetic polypeptide is useful for
 CC modulating immune responses preferably directed against a pathogen or a
 CC cancer. (e.g., cancers of the lung, breast, ovary, cervix, colon, head
 CC and neck, prostate, stomach, bladder, kidney, bone liver,
 CC oesophagus, brain, testicle, uterus), as potentiating agents.
 CC Compositions comprising the polypeptide may be used in the treatment or
 CC prophylaxis against viral (such as infections caused by HIV (human
 CC immunodeficiency virus), hepatitis, influenza, Japanese encephalitis
 CC virus, Epstein-Barr virus and respiratory syncytial virus), bacterial
 CC (e.g., infections caused by Neisseria, Meningococcal, Haemophilus,
 CC Salmonella, Streptococcal, Legionella and Mycobacterium or parasitic
 CC (e.g., infections caused by Plasmodium, Schistosoma, Leishmania,
 CC Trypanosoma, Toxoplasma and Giardia) infections. The present sequence is
 CC a consensus sequence for a parent protein used to design a savine of the
 CC invention.

XX SQ Sequence 307 AA;
 XX Query Match 100.0%; Score 49; DB 6; Length 307;
 Best Local Similarity 100.0%; Pred. No. 2.7;
 Matches 9; Conservative 0; Mismatches 0; Gaps 0;
 YY QY 1 STAPPVHN 9
 Db 170 STAPPVHN 178

XX SQ Sequence 312 AA;
 XX Query Match 100.0%; Score 49; DB 5; Length 312;
 Best Local Similarity 100.0%; Pred. No. 2.8;
 Matches 9; Conservative 0; Mismatches 0; Gaps 0;

XX YY QY 1 STAPPVHN 9
 Db 7 STAPPVHN 15

XX SQ Sequence 312 AA;
 XX Query Match 100.0%; Score 49; DB 5; Length 312;
 Best Local Similarity 100.0%; Pred. No. 2.8;
 Matches 9; Conservative 0; Mismatches 0; Gaps 0;

XX YY QY 1 STAPPVHN 9
 Db 7 STAPPVHN 15

RESULT 12
 AAR96298
 ID AAR96298 standard; peptide; 327 AA.
 XX
 AC AAR96298;
 XX DT 26-JUL-1996 (first entry)
 DE Glycoprotein 39 C terminal fragment.
 XX KW Glycoprotein 39; gp39; lambda gtl1 cDNA library; gastric cancer;
 KW cell line KATO-II; tumour; immune abnormality; marker;
 KW inflammatory disease.
 XX OS Homo sapiens.
 XX PN JP07051065-A.
 XX PD 28-FEB-1995.
 XX PP 21-FEB-1992; 92JP-00035085.
 XX PR 21-FEB-1992; 92JP-00035085.
 XX PA (NIKKEI) NIPPON KOTAI KENKYUSHO KK.
 PA (UTKA-) UNIV KAGOSHIMA.
 XX DR WPI; 1995-127356/17.

RESULT 11
 AAU84810
 ID AAU84810 standard; protein; 312 AA.
 XX
 AC AAU84810;
 XX DT 08-MAY-2002 (first entry)
 DE Human MUC1R consensus sequence.
 XX
 YY Savine; vaccine; cancer; viral infection; HIV; hepatitis C virus;
 KW viral infection; human immunodeficient virus; melanoma;
 KW bacterial infection; Salmonella; Legionella; parasitic infection;
 KW Trypanosoma; Toxoplasma; Giardia.
 XX
 OS Homo sapiens.
 XX PN WO20010197-A1.
 XX PD 29-NOV-2001.
 XX PR 25-MAY-2001; 2001WO-AU000622.
 XX PR 26-MAY-2000; 2000AU-00007761.
 XX PA (AUSU) UNIV AUSTRALIAN NAT.
 XX FI Thomson SA, Ramshaw IA;

DR	N-PSDB; AAT29190.	XX	Chambon P, Kieny MP, Lathe R, Harcavani M;	
PT	Glycoprotein 39 gene - used in the mass production of glyco:protein 39,	PI	WPI; 1992-167097/20.	
PT	PT for use as tumour and immune abnormality marker.	XX	DR N-PSDB; AAQ24679.	
XX	Claim 2; Page 9-10; 10pp; Japanese.	XX	PT Compns. contng. polypeptide antigen recognised by antibody H23 - for treatment of mammary tumors, also for pox virus compns. for use in PR vaccines.	
XX	The sequences given in AAR96297-98 represent portions of glycoprotein 39.	PT	PT treatment of mammary tumors, also for pox virus compns. for use in PR vaccines.	
CC	This sequence represents the C terminal portion of the glycoprotein 39 gene. The cDNA encoding this sequence was isolated from Lambda gt11 cDNA library derived from the gastric cancer cell line KATO-III. Glycoprotein 39 is reported to be used as a tumour marker, an immune abnormality marker or a marker for various inflammatory diseases	XX	XX	
XX	Sequence 327 AA;	XX	The tumour antigen recognised by antibody H23 is aberrantly expressed in epithelial cells from cancerous mammary tissue in about 90 per cent of breast cancer cases; in a normal individual expression is negligible. The antigen exists in two forms: transmembrane (ETA-T) and secreted (ETA-S). Both forms show a high degree of polymorphism. A 20 amino acid subunit in ETA can be tandemly repeated up to 80 times. (N.B. two tandem repeats are shown here; the first half of the amino acid sequence, i.e. on the N-terminal side of the repeat region, is given in AAR2766). From one CC submit to the next, 1 to 3 amino acids can differ. See also AAQ24678- correct PN field.)	
SQ	Sequence 327 AA;	XX	CC	
Query Match	Score 49; DB 2; Length 327;	Query Match	Score 49; DB 2; Length 348;	
Best Local Similarity	100.0%;	Best Local Similarity	100.0%;	
Matches	9;	Matches	9;	
AC	Conservative 0;	AC	Mismatches 0;	
XX	Mismatches 0;	XX	Indels 0;	
Qy	1 STAPPVINV 9	Qy	1 STAPPVINV 9	
Db	22 STAPPVINV 30	Db	43 STAPPVINV 51	
XX	XX	XX	Sequence 348 AA;	
RESULT 13	AAR27662	ID AAR27662 standard; protein; 348 AA.	RESULT 14	
XX	XX	XX	AAR23973 standard; protein; 455 AA.	
AC	AC	AC	XX	
XX	XX	XX	AAR23973;	
XX	XX	XX	XX	
DT	25-MAR-2003 (revised)	DT	25-MAR-2003 (revised)	
DT	06-NOV-1992 (first entry)	DT	06-NOV-1992 (first entry)	
XX	XX	XX	DE Transmembrane form of H23-ETA antigen.	
XX	KW Transmembrane; human epithelial antigen; Monoclonal antibody H23; vaccine; malignant tumour; breast cancer; tandem repeat.	XX	XX	
XX	KW	XX	DE Transmembrane form of H23-ETA antigen.	
XX	C-terminal region of H23-ETA antigen.	XX	XX	
DE	XX	XX	KW ETA-T; human epithelial antigen; Monoclonal antibody H23; vaccine; malignant tumour; breast cancer; tandem repeat.	
XX	Homo sapiens.	XX	XX	
OS	XX	XX	Homo sapiens.	
XX	XX	XX	OS	
FH	Key	Location/Qualifiers	XX	
FT	Region	1..40	PH	
FT	/note= "contains 2 tandem repeats - can have up to 80 copies"	FT	Peptide	
FT	FT	/label= Pro, Ala	FT	FT
FT	/note= "natural polymorphism"	FT	Protein	FT
FT	Misc-difference 7	FT	Misc-difference 134	FT
FT	/label= Pro, Ala	FT	/label= Pro, Ala	FT
FT	/note= "natural polymorphism"	FT	/note= "natural polymorphism"	FT
FT	Misc-difference 17	FT	/note= "natural polymorphism"	FT
FT	/label= Thr, Asn	FT	/note= "natural polymorphism"	FT
FT	/note= "natural polymorphism"	FT	/note= "natural polymorphism"	FT
FT	Misc-difference 20	FT	/note= "natural polymorphism"	FT
FT	/label= Pro, Ala	FT	/note= "natural polymorphism"	FT
FT	/note= "natural polymorphism"	FT	/note= "natural polymorphism"	FT
FT	Misc-difference 27	FT	/note= "natural polymorphism"	FT
FT	/label= Pro, Ala	FT	/note= "natural polymorphism"	FT
FT	/note= "natural polymorphism"	FT	/note= "natural polymorphism"	FT
FT	Misc-difference 37	FT	/note= "natural polymorphism"	FT
FT	/label= Thr, Asn	FT	/note= "natural polymorphism"	FT
FT	/note= "natural polymorphism"	FT	/note= "natural polymorphism"	FT
FT	Misc-difference 40	FT	/note= "natural polymorphism"	FT
FT	/label= Pro, Ala	FT	/note= "natural polymorphism"	FT
FT	/note= "natural polymorphism"	FT	/note= "natural polymorphism"	FT
XX	XX	XX	Misc-difference 147	FT
PN	WO9207000-A1.	PN	/label= Pro, Ala	FT
XX	XX	XX	/note= "natural polymorphism"	FT
PD	30-APR-1992.	XX	XX	FT
XX	XX	XX	WO9207000-A1.	FT
PF	23-OCT-1991;	PD	30-APR-1992.	FT
PR	23-OCT-1990;	XX	XX	FT
XX	90FR-00013101.	XX	23-OCT-1991;	FT
PA	(TRGE) TRANSGENE SA.	XX	91WO-FR0000835.	FT
PA	23-OCT-1990;	XX	23-OCT-1990;	FT
PA	90FR-00013101.	XX	90FR-00013101.	FT

XX (TRGE) TRANSGENE SA,
 PA PT Compens. contng. Polypeptide antigen recognised by antibody H23 - for
 XX Chambon P, Kiery MP, Lathe R, Hareveni M;
 PI PT treatment of mammary tumours, also for pox virus compsns. for use in
 XX PT vaccines.
 XX PS Claim 3; Page 15-18; 29pp; French.
 XX The tumour antigen recognised by antibody H23 is aberrantly expressed in
 CC epithelial cells from cancerous mammary tissue in about 90 per cent of
 CC breast cancer cases; in a normal individual expression is negligible. The
 CC antigen exists in two forms: transmembrane (ETA-T) and secreted (ETA-S).
 CC Both forms show a high degree of polymorphism. A 20 amino acid subunit in
 CC ETA can be tandemly repeated up to 80 times. From one subunit to the
 CC next, 1 to 3 amino acids can differ. DNA coding for immunogenic fragments
 CC of ETA can be inserted into e.g. vaccinia viruses for treatment of
 CC mammary tumours. See also AAQ24478-Q24681, AAQ29277 and AAR23974-R23981.
 CC (Updated on 25-MAR-2003 to correct PN field.)

XX SQ Sequence 455 AA;

XX Query Match 100.0%; Score 49; DB 2; Length 455;
 Best Local Similarity 100.0%; Pred. No. 4;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX QY 1 STAPPVHN 9
 Db 150 STAPPVHN 158

XX SQ Sequence 455 AA;

XX Query Match 100.0%; Score 49; DB 2; Length 455;
 Best Local Similarity 100.0%; Pred. No. 4;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX QY 1 STAPPVHN 9
 Db 150 STAPPVHN 158

RESULT 15
 AAY71024
 ID AAY71024 standard; protein; 455 AA.
 XX DT 29-AUG-2000 (first entry)
 XX DE Human Mucin 1 (MUC-1) protein fragment #5.
 XX KW Human; Mucin 1; MUC-1; tumour; PMRS30 expression vector; anti-tumour,
 KW therapy; immune response; cytostatic; vaccine.
 XX OS Homo sapiens .
 XX PN WO2000025827-A2.
 XX PD 11-MAY-2000.
 XX PF 18-OCT-1999; 99WO-EP007874.
 XX PR 30-OCT-1998; 98ITI-MI002330.
 XX PA (MENA) MENARINI RICERCHE SPA.
 XX Parente D, Di Massimo AM, De Santis R;
 XX DR WPI; 2000-365410/31.
 XX DR N-PSDB; AAD0388.

XX PT Composition containing one or more DNA molecules encoding fragments of a
 PT Mucin 1 (MUC-1) protein overexpressed in tumor cells, useful in anti-
 PT tumor therapy.

XX Claim 16; Fig 5; 56pp; English.

CC The present sequence is a fragment of human Mucin 1 (MUC-1), an antigenic